1-1-2005

NMR refinement of under-determined loop regions of the E200K variant of the human prion protein using database-derived distance constraints

Kriti Gautam Mukhopadhyay
Iowa State University

NMR Refinement of under-determined loop regions of the E200K variant of the human prion protein using database-derived distance constraints

by

Kriti Gautam Mukhopadhyay

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Applied Mathematics

Program of Study Committee:
Zhijun Wu, Major Professor
Maria Axenovich
Paul Sacks
Won-BinYoung

Iowa State University
Ames, Iowa
2005

Graduate College
Iowa State University

This is to certify that the master's thesis of

Kriti Gautam Mukhopadhyay

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF GRAPHS

ABSTRACT

Computational studies and research conducted in order to facilitate the understanding of the conversion of the normal cellular prion (PrP^C) to the scrapie prion (PrP^{Sc}) in prion diseases, are usually based on the structures determined by NMR. This is mainly attributed to the difficulties involved in crystallizing the prion protein. Due to insufficient experimental restraints, a biologically critical loop region in PrP^c (residues 167-171), which is the potential binding site for the hypothized Protein X, is under-determined in most mammalian species. In this research, we show that by adding information about distance constraints derived from a database of high-resolution protein structures, this under-determined loop and some other secondary structural elements of the E200K variant of human PrP^C can be refined into more generally realistic and acceptable structures within an ensemble, with improved quality and increased accuracy.

In particular, the ensemble becomes more compact after the refinement with database derived distances constraints and the percentage of residues in the most favourable region of the Ramachandran diagram is increased to about 90% in the refined structures from the 80 to 85% range in the previously reported structures. In NMR structures, a model with 90% or more residues lying in the most favourable regions of the Ramachandran plot, is considered a good quality model. Our results not only provide a significantly improved model of structures of the Human prion protein, that would hence facilitate insights into its conversion in the spongiform encephalopathies, but also demonstrate the strong potential for using databases of known protein structures for structure determination and refinement.

CHAPTER 1
INTRODUCTION

1.1 Introduction

Transmissible spongiform encephalopathies, or prion diseases, are a group of neurodegenerative diseases in mammalian species, characterized by a progressive vacuolation of brain tissue, amyloid protein deposits, and astrogliosis. These diseases include scrapie in sheep, bovine spongiform encephalopathy (Mad Cow disease) in cattle, and Creutzfeldt-Jakob disease (CJD), Gerstmann-Straüssler- Scheinker disease (GSS), fatal familial insomnia (FFI), and kuru in humans. These diseases may occur sporadically, can be inherited, or may be transmitted.

The pathogenesis of the prion diseases is associated with accumulation of the infectious 'scrapie' form of the prion protein (PrP^{Sc}) in the brain tissue, which is transformed from its normal cellular form (PrP^{C}). Hereditary forms of the disease are linked to specific mutations in the gene coding for the prion protein (PRNP). For instance, the familial form of CJD is linked with a point mutation on PRNP that results in a residue substitution at the 200^{th} residue.

Yet, to date, the mechanism of the $PrP^{C} \rightarrow PrP^{Sc}$ conversion, which is considered the key process in the pathogenesis of prion diseases, remains unclear.

One of the obstacles in understanding the details of this conformational conversion is that the PrP^{Sc} sample is hard to purify for biochemical and structural characterization. The cellular and scrapie isoforms of PrP have also proven difficult for high-resolution spectroscopic or crystallographic study. Therefore, high-quality structures of the protein are urgently needed to provide better insight into its transition process.

Thus far, only two X-ray structures of PrP^{C} have been reported. This is due to the fact that like other membrane glycoproteins, the prion protein is extremely difficult to crystallize when glycosylated. Most normal and disease-related variants of PrP^{C} have been determined by Nuclear Magnetic resonance spectroscopy (NMR). The solution structure of the E200K variant of the Human prion protein, which is associated with familial CJD, was determined by multi-dimensional NMR spectroscopy by Zhang et al.

Due to the lack of NOE (Nuclear Over Hauser Effects) restraints, one particular loop region that comprises residues 167-171, is under-determined in these NMR structures. This region is well-conserved in

mammalian species and is believed to be the binding site for the hypothized chaperone in the transition process – 'Protein X'. A structure quality analysis by PROCHECK shows that none of the residues in this critical region fall in the most favorable regions of the Ramachandran plots. This suggests that it may be highly significant to refine the structure in this region to elucidate the interaction between the prion protein and Protein X. An alternative explanation could be that this loop is flexible, and the NMR structure is built from data reflecting some average of these forms that might not actually be a feasible form.

The refinement of the NMR structures and the loop region mentioned above can be achieved by implementing standard peptide information, such as dihedral angles and inter-atomic distances, based on statistical analysis of databases of high-resolution protein structures. In particular, it has been shown that inter-atomic distance constraints can improve NMR structures yielding increased precision and accuracy, without compromising the quality of the NMR structures. Moreover, these constraints impose literally no extra cost on the NMR structure refinement.

In this research, we used a specific set of inter-atomic distance constraints between heavy atoms derived from a database of high-resolution protein structures. We used these constraints as additional constraints in the NMR refinement process, to refine the E200K variant of the Human prion protein.

Our results show that the critical loop region (residues 167-171), as well as some other secondary structure elements of the protein, were significantly improved in terms of precision and accuracy, and the Ramachandran plots of the structures, when the additional constraints are implemented. This is the first evidence that these distance constraints can be used to optimize the under-determined regions of a protein (in this case, the human prion protein). The results provide significantly improved structural information about the prion protein and hence could ultimately provide a better insight in the conversion process in the pathogenesis of prion diseases.

In general, it can be expected that this approach will be highly valued in refinement of under-defined NMR structures.

1.2 Terms and Abbreviations

PrP^C	Normal cellular form of the Prion Protein
1FO7	The PDB ID for the E200K variant of the Human Prion Protein
CJD	Creutzfeldt Jakob Disease, a Prion disease
FTIR	Fourier Transform Infra-red spectroscopy, used to analyze proteins
GSS	Gerstmann- Straüssler- Scheinker Disease, a Prion disease
HuPrP	Human Prion Protein
IFT	Inverse Fourier Transform
Kuru	Prion disease observed in Papa New Guinea amongst the Foré- speaking people
MoPrP	Mouse Prion protein
NMR	Nuclear Magnetic Resonance spectroscopy, used to analyze protein structures
NOE	Nuclear Over Hauser Effect
PDB	Protein Data Bank
PRNP	Gene encoding PrP
PrP^{Sc}	Infectious scrapie form of the Prion Protein
RMSD	Root mean square deviation
SA	Simulated Annealing
SHaPrP	Syrian Hamster Prion Protein
shPrP	Sheep Prion Protein
Tg	Transgenic, i.e., with foreign DNA inserted in the genome
Transgene	Foreign DNA expressed in a 'transgenic' animal
TSE	Transmissible Spongiform Encephalopathy, generalized term for prion diseases

CHAPTER 2.

LITERATURE REVIEW

2.1 Basics of NMR

NMR

Nuclear Magnetic Resonance (NMR) is a phenomenon that occurs when the nuclei of certain atoms are immersed in a static magnetic field and exposed to another oscillating magnetic field. Certain nuclei experience this phenomenon while others don't, depending on whether they possess a property called 'spin'.

Spectroscopy is the study of the interaction of electromagnetic radiation with matter. NMR spectroscopy is the use of the NMR phenomenon to study physical, chemical and biological properties of matter. It is used by chemists to study chemical structure using simple one-dimensional techniques. Two-dimensional techniques are used to probe molecular dynamics in solutions. Solid state NMR spectroscopy is used to determine the molecular structure of solids. Other scientists have developed NMR methods of measuring diffusion coefficients.

Spin Physics

Spin is a fundamental property of nature like electrical charge or mass. It comes in multiples of ½ and can be + or -. Protons, electrons, and neutrons possess spin. Individual unpaired electrons, protons, and neutrons each possess a spin of ½.

Two or more particles with spin having opposite signs can pair up to eliminate the observable manifestations of spin. In NMR, it is unpaired nuclear spins that are of importance.

When placed in a magnetic field of strength B, a particle with a net spin can absorb a photon, of frequency ν. This frequency depends on the gyro magnetic ratio γ of the particle.

$$\nu = \gamma B$$

To understand how particles behave in a magnetic field, consider a proton. This proton has the property called spin. Think of the spin as a magnetic moment vector, causing the proton to behave like a tiny magnet with north and south poles. When the proton is placed in an external magnetic field, the spin vector aligns itself with the external field, similar to a magnet. There is a low energy configuration (a state where the poles are aligned N-S-N-S) and a high energy configuration (a state where poles are aligned N-N-S-S). This particle can undergo a transition between the two energy states by the absorption of a photon. A particle in the lower energy state can end up in the higher energy state by absorbing a photon. The energy of this photon must exactly match the energy difference between the two states. The energy, E, of the photon, is related to its frequency, ν, by Planck's constant ($h = 6.626 * 10^{-34}$ J s).

$$E = h \nu$$

In NMR and MRI(Magnetic Resonance Imaging), the quantity ν is called the resonance frequency and the Larmor frequency.

Therefore, we get, the energy of the photon needed to cause a transition between the two spin states is

$$E = h \nu B$$

When the energy of the photon matches the energy difference between the two spin states, absorption of energy occurs.

In NMR experiments, the frequency of the photon is in the radio frequency (RF) range. In NMR spectroscopy, ν is between 60 and 800 MHz for hydrogen nuclei.

The simplest NMR experiment is the continuous wave experiment. There are two ways of performing it:

(i) A continuous constant frequency probes the energy level, while the magnetic field is varied.

(ii) Under a constant magnetic field, while the frequency is varied.

At room temperature, the number of spins in the lower energy level, N^+, slightly outnumbers the number in the higher energy level, N^-. Boltzmann statistics reveals that

$$N^-/N^+ = e^{-E/kT}$$

where E = energy difference between spin states

k = Boltzmann constant with value $1.3805 * 10^{-23}$ J/Kelvin

T = temperature in Kelvin

As T decreases, N^-/N^+ decreases. As T increases, $N^-/N^+ \rightarrow 1$

The signal in NMR spectroscopy results from the difference between the energy absorbed by the spins (which make a transition from lower energy state to the higher energy state), and the energy emitted by the spins (which simultaneously make a transition from higher energy state to the lower energy state). Thus, the signal is proportional to the population difference between the two states. It is the resonance, or exchange of energy at a specific frequency between the spins and the spectrometer, which gives NMR its sensitivity.

To describe NMR on a macroscopic scale, one needs to define spin packets. A spin packet is a group of spins experiencing the same magnetic field strength. At any instant in time, the magnetic field due to the spins in each spin packet can be represented by a magnetization vector. The size of each vector is proportional to $(N^+ - N^-)$. The vector sum of the magnetization vectors from all the spin packets is the net magnetization. Adapting the conventional NMR coordinate system, the external magnetic field and the net magnetization vector at equilibrium are both along the Z-axis.

At equilibrium, the net magnetization vector lies along the direction of the applied magnetic field B_0 and is called the equilibrium magnetization M_0. In this configuration, the Z component of magnetization M_Z equals M_0. M_Z is referred to as the longitudinal magnetization. There is no transverse magnetization (X and Y components) here. One can change the net magnetization by exposing the nuclear spin system to energy of a

frequency equal to the energy difference between the spin states. If enough energy is put into the system, it is possible to saturate the spin system and make $M_Z = 0$.

The time constant that describes how M_Z returns to its equilibrium value, is called the spin lattice relaxation time, denoted by T_1. The equation governing this behavior as a function of the time t after its displacement is :

$$M_Z = M_0 (1 - 2e^{-t/T1})$$

T_1 is the time to reduce the difference between the longitudinal magnetization and its equilibrium value by a factor of e.

If the net magnetization is placed in the XY plane, it will rotate about the Z-axis at a frequency equal to the frequency of the photon which would case a transition between the two energy levels of the spin. This frequency is called the Larmor frequency.

In addition to the rotation, the net magnetization starts to diphase because each of the spin packets making it up is experiencing a slightly different magnetic field and rotates at its own Larmor frequency. The longer the elapsed time, the greater the phase difference.

The time constant which describes the return to equilibrium of the transverse magnetization (M_{XY}), is called the spin-spin relaxation time, denoted by T_2.

$$M_{XY} = M_{XY0}\, e^{-t/T2}$$

T_2 = time to reduce the transverse magnetization by a factor of e. It is always less than or equal to the spin lattice relaxation time.

The combination of two factors contribute to the decay of transverse magnetization:

a) molecular interactions (pure T_2 molecular effect)

b) variations in B_0 (inhomogeneous T_2 effect)

The combined time constant is denoted by T_2^*. The relationship between these constants is given by:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_{2inh}}$$ where T_{2inh} := inhomogeneous T_2 effect

It is convenient to define a rotating frame of reference which rotates about the Z-axis at the Larmor frequency. We distinguish this coordinate system from the laboratory system by X'Y'. A transverse magnetization vector rotating about the Z axis at the same velocity as the rotating frame will appear stationary in the rotating frame. A magnetization vector traveling faster than the rotating frame rotates clockwise about the Z axis. A magnetization vector traveling slower than the rotating frame rotates counter-clockwise about the Z axis.

A coil of wire placed around the X axis will provide a magnetic field along the X axis when a direct current is passed through the coil. An alternating current will produce a magnetic field which alternates in direction. In a frame of reference rotating about the Z axis at a frequency equal to that of the alternating current, the magnetic field along the X' axis will be constant. This is the same as moving the coil about the rotating frame coordinate system at the Larmor Frequency. In magnetic resonance, the magnetic field created by the coil passing an alternating current at the Larmor frequency is called the B_1 magnetic field. When the alternating current through the coil is turned on and off, it creates a pulsed B_1 magnetic field along the X' axis. The spins respond to this pulse in such a way as to cause the net magnetization vector to rotate about the direction of the applied B_1 field. The rotation angle depends on the length of time the field is on, τ, and its magnitude B_1.

$$\theta = 2 \pi \gamma \tau B_1$$

A 90^o pulse is one which rotates the magnetization vector clockwise by 90 degrees about the X' axis. A 90^o pulse rotates the equilibrium magnetization down to the Y' axis. In the laboratory frame the equilibrium magnetization spirals down around the Z axis to the XY plane. An 180^o pulse will rotate the magnetization vector by 180 degrees. An 180^o pulse rotates the equilibrium magnetization down to along the -Z axis. You can see why the rotating frame of reference is helpful in describing the behavior of magnetization in response to a pulsed magnetic field. The net magnetization at any orientation will behave according to the rotation equation. For example, a net magnetization vector along the Y' axis will end up along the -Y' axis when acted upon by an 180^o pulse of B_1 along the X' axis. A net magnetization vector between X' and Y' will end up between X' and Y' after the application of an 180^o pulse of B_1 applied along the X' axis.

A rotation matrix can also be used to predict the result of a rotation. Here θ is the rotation angle about the X' axis, [*X'*, *Y'*, *Z*] is the initial location of the vector, and [*X''*, *Y''*, *Z''*] the location of the vector after the rotation.

$$\begin{bmatrix} X'' \\ Y'' \\ Z'' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix}$$

Motions in solution which result in time varying magnetic fields cause spin relaxation. Time varying fields at the Larmor frequency cause transitions between the spin states and hence a change in M_Z.

There is a distribution of rotation frequencies in a sample of molecules. Only frequencies at the Larmor frequency affect T_1. Since the Larmor frequency is proportional to B_o, T_1 will therefore vary as a function of magnetic field strength. In general, T_1 is inversely proportional to the density of molecular motions at the Larmor frequency.

The rotation frequency distribution depends on the temperature and viscosity of the solution. Therefore T_1 will vary as a function of temperature. At the Larmor frequency indicated by ν_o, T_1 (280 K) < T_1 (340 K). The temperature of the human body does not vary by enough to cause a significant influence on T_1. The viscosity does however vary significantly from tissue to tissue and influences T_1 as is seen in Figure 2.1.

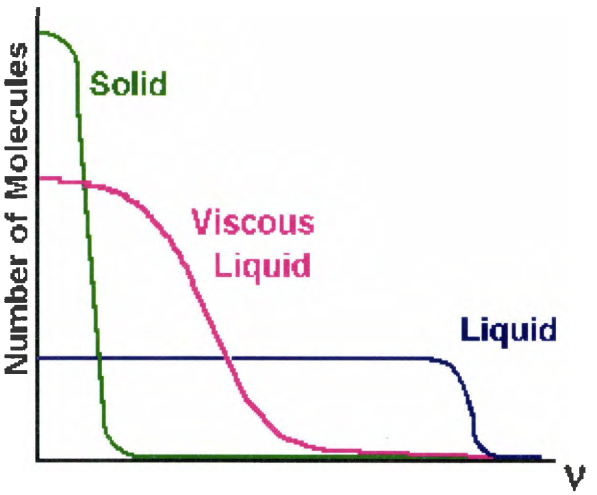


Figure 2.1. Molecular motion plot

Source: http://www.cis.rit.edu/htbooks/nmr

Fluctuating fields which perturb the energy levels of the spin states cause the transverse magnetization to dephase. In general, relaxation times get longer as B_o increases because there are fewer relaxation-causing frequency components present in the random motions of the molecules.

Exchanges

Spin exchange is the exchange of spin state between two spins. For example, if we have two spins, A and B, and A is spin up and B is spin down, spin exchange between A and B can be represented with the following equation.

$$A(\uparrow) + B(\downarrow) \quad \leftrightarrow \quad A(\downarrow) + B(\uparrow)$$

The bidirectional arrow indicates that the exchange reaction is reversible.

The energy difference between the upper and lower energy states of A and of B must be the same for spin exchange to occur. On a microscopic scale, the spin in the upper energy state (B) is emitting a photon which is being absorbed by the spin in the lower energy state (A). Therefore, B ends up in the lower energy

state and A in the upper state. Spin exchange will not affect T_1 but will affect T_2. T_1 is not affected because the distribution of spins between the upper and lower states is not changed. T_2 will be affected because phase coherence of the transverse magnetization is lost during exchange.

Another form of exchange is called chemical exchange. In chemical exchange, the A and B nuclei are from different molecules. Consider the chemical exchange between water and ethanol.

$$CH_3CH_2OH_A + HOH_B \leftrightarrow CH_3CH_2OH_B + HOH_A$$

Here the B hydrogen of water ends up on ethanol, and the A hydrogen on ethanol ends up on water in the forward reaction. There are four scenarios for the nuclear spin, represented by the four equations.

$$CH_3CH_2OH_A(\uparrow) + HOH_B(\downarrow) \leftrightarrow CH_3CH_2OH_B(\downarrow) + HOH_A(\uparrow)$$

$$CH_3CH_2OH_A(\downarrow) + HOH_B(\uparrow) \leftrightarrow CH_3CH_2OH_B(\uparrow) + HOH_A(\downarrow)$$

$$CH_3CH_2OH_A(\uparrow) + HOH_B(\uparrow) \leftrightarrow CH_3CH_2OH_B(\uparrow) + HOH_A(\uparrow)$$

$$CH_3CH_2OH_A(\downarrow) + HOH_B(\downarrow) \leftrightarrow CH_3CH_2OH_B(\downarrow) + HOH_A(\downarrow)$$

Chemical exchange will affect both T_1 and T_2. T_1 is now affected because energy is transferred from one nucleus to another. For example, if there are more nuclei in the upper state of A, and a normal Boltzmann distribution in B, exchange will force the excess energy from A into B. The effect will make T_1 appear smaller. T_2 is affected because phase coherence of the transverse magnetization is not preserved during chemical exchange.

Bloch Equations

The Bloch equations are a set of coupled differential equations which can be used to describe the behavior of a magnetization vector under any conditions.

$$\frac{dM_{x'}}{dt} = (\omega_0 - \omega) M_{y'} - \frac{M_{x'}}{T_2}$$

$$\frac{dM_{y'}}{dt} = -(\omega_0 - \omega) M_{x'} + 2\pi\gamma B_1 M_z - \frac{M_{y'}}{T_2}$$

$$\frac{dM_z}{dt} = -2\pi\gamma B_1 M_{y'} - \frac{(M_z - M_{z0})}{T_1}$$

When properly integrated, the Bloch equations will yield the X', Y', and Z components of magnetization as a function of time.

Chemical Shift

When an atom is placed in a magnetic field, its electrons circulate about the direction of the applied magnetic field. This circulation causes a small magnetic field at the nucleus which opposes the externally applied field. The magnetic field at the nucleus (the effective field) is therefore generally less than the applied field by a fraction σ.

$$B = B_o (1 - \sigma)$$

In some cases, such as the benzene molecule, the circulation of the electrons in the aromatic π orbitals creates a magnetic field at the hydrogen nuclei which enhances the B_o field. This phenomenon is called de-shielding. The electron density around each nucleus in a molecule varies according to the types of nuclei and bonds in the molecule. The opposing field and therefore the effective field at each nucleus will vary. This is called the *chemical shift phenomenon.*

The chemical shift of a nucleus is the difference between the resonance frequency of the nucleus and a standard, relative to the standard. This quantity is reported in ppm and given the symbol delta, δ.

$$\delta = \frac{(\nu - \nu_{REF}) * 10^6}{\nu_{REF}}$$

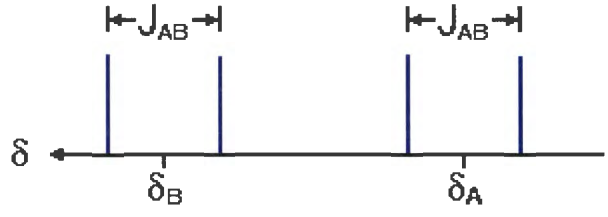
In NMR spectroscopy, this standard is often tetramethylsilane, $Si(CH_3)_4$, abbreviated TMS. The chemical shift is a very precise metric of the chemical environment around a nucleus. The magnitude of the screening depends on the atom.

J-Coupling

Nuclei experiencing the same chemical environment or chemical shift are called *equivalent*. Those nuclei experiencing different environment or having different chemical shifts are *nonequivalent*. Nuclei which are close to one another exert an influence on each other's effective magnetic field. This effect shows up in the NMR spectrum when the nuclei are nonequivalent. If the distance between non-equivalent nuclei is less than or equal to three bond lengths, this effect is observable. This effect is called *spin-spin coupling* or *J- coupling*.

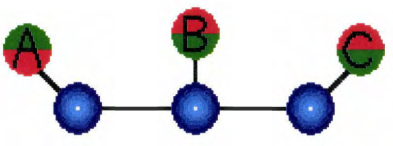
The absorption frequencies are represented by vertical lines between the energy levels in the NMR spectrum (energy level diagram). The distance between two split absorption lines is called the J-coupling constant or the spin-spin splitting constant and is a measure of the magnetic interaction between the two nuclei. Energy levels with the same energy are said to be degenerate.

Figure 2.1 (a)Illustration of the NMR spectrum for two nuclei, A and B, reflecting the splitting observed in the energy level diagram.



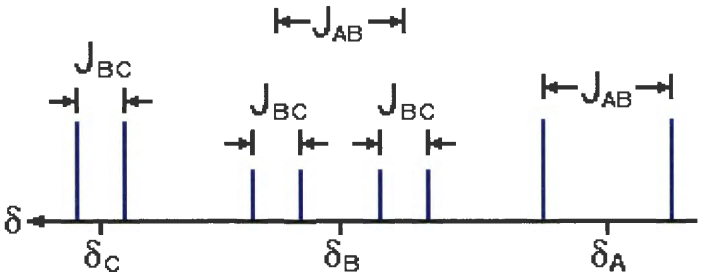
When there are two different types of nuclei three bonds away, there will be two values of J, one for each pair of nuclei. For instance if we have three nuclei, A,B and C as shown:

Figure 2.1.(b)



then, the energy level diagram would be represented as follows:

Figure 2.1 (c) Energy level diagram



In the above example, $J_{AB} > J_{BC}$.

Time Domain NMR Signal

An NMR sample may contain many different magnetization components, each with its own Larmor frequency. These magnetization components are associated with the nuclear spin configurations joined by an allowed transition line in the energy level diagram. Based on the number of allowed absorptions due to chemical shifts and spin-spin couplings of the different nuclei in a molecule, an NMR spectrum may contain many different frequency lines.

In pulsed NMR spectroscopy, signal is detected after these magnetization vectors are rotated into the XY plane. Once a magnetization vector is in the XY plane it rotates about the direction of the B_0 field, the +Z axis. As transverse magnetization rotates about the Z axis, it will induce a current in a coil of wire located

around the X axis. Plotting current as a function of time gives a sine wave. This wave will, of course, decay with time constant T_2^* due to dephasing of the spin packets. This signal is called a *free induction decay* (FID).

By convention, transverse magnetization vectors rotating faster than the rotating frame of reference are said to be rotating at a positive frequency relative to the rotating frame (+ν). Vectors rotating slower than the rotating frame are said to be rotating at a negative frequency relative to the rotating frame (-ν).

In most NMR spectra, the resonance frequency of a nucleus, as well as the magnetic field experienced by the nucleus and the chemical shift of a nucleus, increase from right to left.

Fourier Transforms

A Fourier transform is an operation which converts functions from time to frequency domains. An inverse Fourier transform (IFT) converts from the frequency domain to the time domain.

A magnetization vector, starting at +x, is rotating about the Z axis in a clockwise direction. The plot of M_x as a function of time is a cosine wave. Fourier transforming this gives peaks at both +ν and - ν because the Fourier Transform can not distinguish between a + ν and a - ν rotation of the vector from the data supplied.

Similarly, a plot of M_y as a function of time is a -sine function. Fourier transforming this gives peaks at + ν and – ν.

The solution is to input both the M_x and M_y into the Fourier Transform. The Fourier transform is designed to handle two orthogonal input functions called the real and imaginary components. Detecting just the M_x or M_y component for input into the Fourier transform is called *linear detection*. This was the detection scheme on many older NMR spectrometers and some magnetic resonance imagers. Detection of both M_x and M_y is called *quadrature detection* and is the method of detection on modern spectrometers and imagers. It is the method of choice since now the Fourier transform can distinguish between + ν and - ν, and all of the frequency domain data be used.

A Fourier Transform is defined by the integral

$$f(\omega) = \int_{-\infty}^{+\infty} f(t)\, e^{-i\omega t}\, dt = \int_{-\infty}^{+\infty} f(t)\left[\cos(\omega t) - i\sin(\omega t)\right] dt$$

Think of f(ω) as the overlap of f(t) with a wave of frequency ω.

$$f(\omega) = \sum_{-\infty}^{+\infty} f(t)\left[\cos(\omega t) - i\sin(\omega t)\right]$$

This is easy to picture by looking at the real part of $f(\omega)$ only.

$$f(\omega) = \sum_{-\infty}^{+\infty} f(t)\cos(\omega t)$$

Consider the function of time, f(t) = cos(4t) + cos(9t).

The inverse Fourier transform (IFT) is best depicted as an summation of the time domain spectra of frequencies in f(ω).

Phase Correction

The actual Fourier transform will make use of an input consisting of a real and an imaginary part. You can think of M_x as the real input, and M_y as the imaginary input. The resultant output of the transform will therefore have a real and an imaginary component, too.

Consider the following function:

$$f(t) = e^{-at}\, e^{-i2\pi\nu t}$$

In NMR spectroscopy, the real output of the Fourier Transform is taken as the frequency domain spectrum. To see an esthetically pleasing (absorption) frequency domain spectrum, we want to input a cosine function into the real part and a sine function into the imaginary parts of the FT. This is what happens if the cosine part is input as the imaginary and the sine as the real.

To obtain an absorption spectrum as the real output of the transform, a phase correction must be applied to either the time or frequency domain spectra. This process is equivalent to the coordinate transformation described earlier

$$\begin{bmatrix} RE'' \\ IM'' \end{bmatrix} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} RE \\ IM \end{bmatrix}$$

If the above mentioned FID is recorded such that there is a 40^{o} phase shift in the real and imaginary FIDs, the coordinate transformation matrix can be used with $\varphi = -45^{o}$. The corrected FIDs look like a cosine function in the real and a sine in the imaginary.

Fourier transforming the phase corrected FIDs gives an absorption spectrum for the real output of the FT. This correction can be done in the frequency domain as well as in the time domain.

NMR spectra require both constant and linear corrections to the phasing of the Fourier transformed signal.

$$\Phi = m\,\nu + b$$

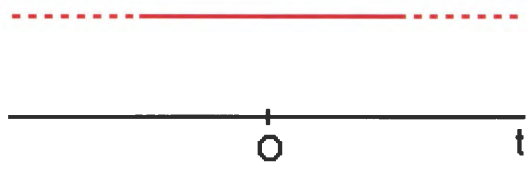
Constant phase corrections, b, arise from the inability of the spectrometer to detect the exact M_x and M_y. Linear phase corrections, m, arise from the inability of the spectrometer to detect transverse magnetization starting immediately after the RF pulse.

In magnetic resonance, the M_x or M_y signals are displayed. A magnitude signal might occasionally be used in some applications. The magnitude signal is equal to the square root of the sum of the squares of M_x and M_y.

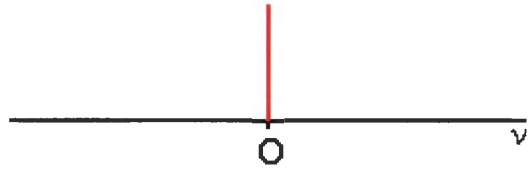
To better understand Fourier Transform NMR functions, you need to know some common Fourier pairs. A Fourier pair is two functions, the frequency domain form and the corresponding time domain form. Here are a few Fourier pairs which are useful in NMR (Source: http://www.cis.rit.edu/htbooks/nmr). The amplitude of the Fourier pairs has been neglected since it is not relevant in NMR. FT denotes the Fourier Transform.

- Constant value at all time :

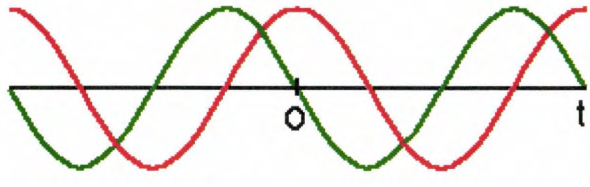
A DC offset or constant value.



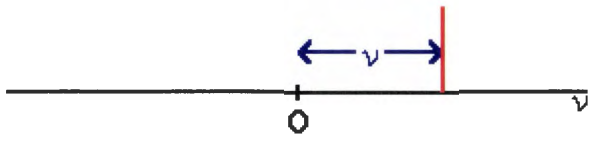
A delta function at zero.



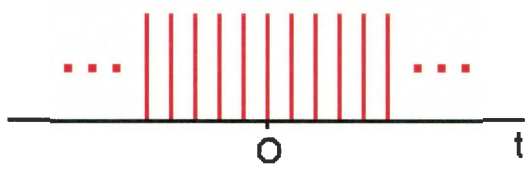
- Real: $\cos(2\pi\nu t)$, Imaginary: $-\sin(2\pi\nu t)$



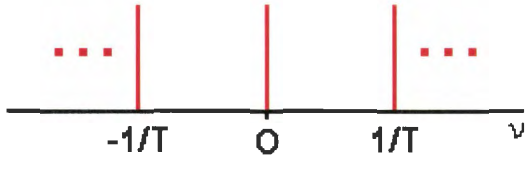
A delta function at π.



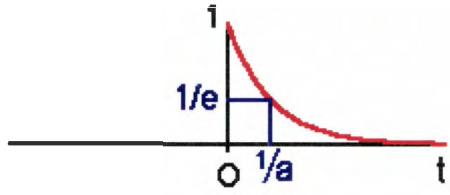
- Comb Function (A series of delta functions separated by T.)



A comb function with separation 1/T.



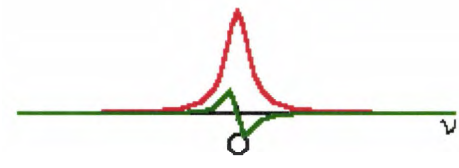
- Exponential Decay: e^{-at} for $t > 0$.



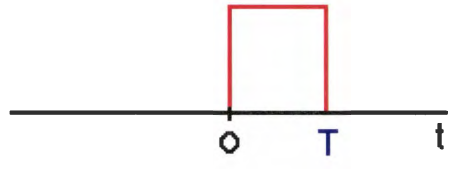
Lorentzian

$$\text{RE: } \frac{a^2}{a^2 + 4\pi^2\nu^2}$$

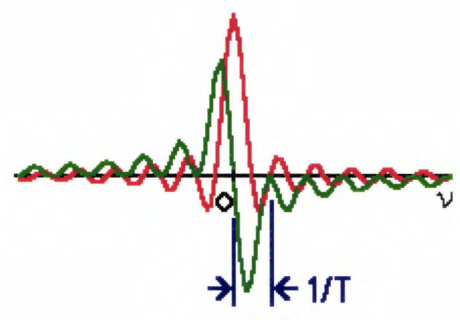
$$\text{IM: } \frac{2a^2\pi\nu}{(a^2 + 4\pi^2\nu^2)^2}$$



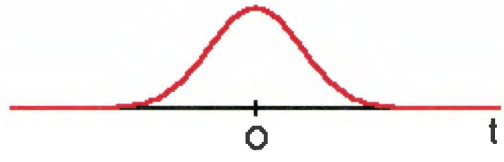
- A square pulse starting at 0 that is T seconds long.



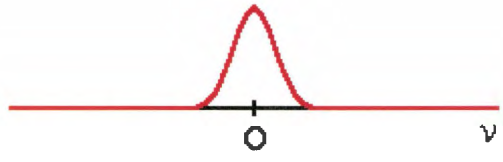
Sinc RE: $\frac{\sin(2\pi\nu t)}{(2\pi\nu t)}$ IM: $\frac{-(\sin^2(2\pi\nu t))}{\pi\nu t}$



- Gaussian: e^{-at^2}



Gaussian: $e^{-\pi^2\nu^2/a}$

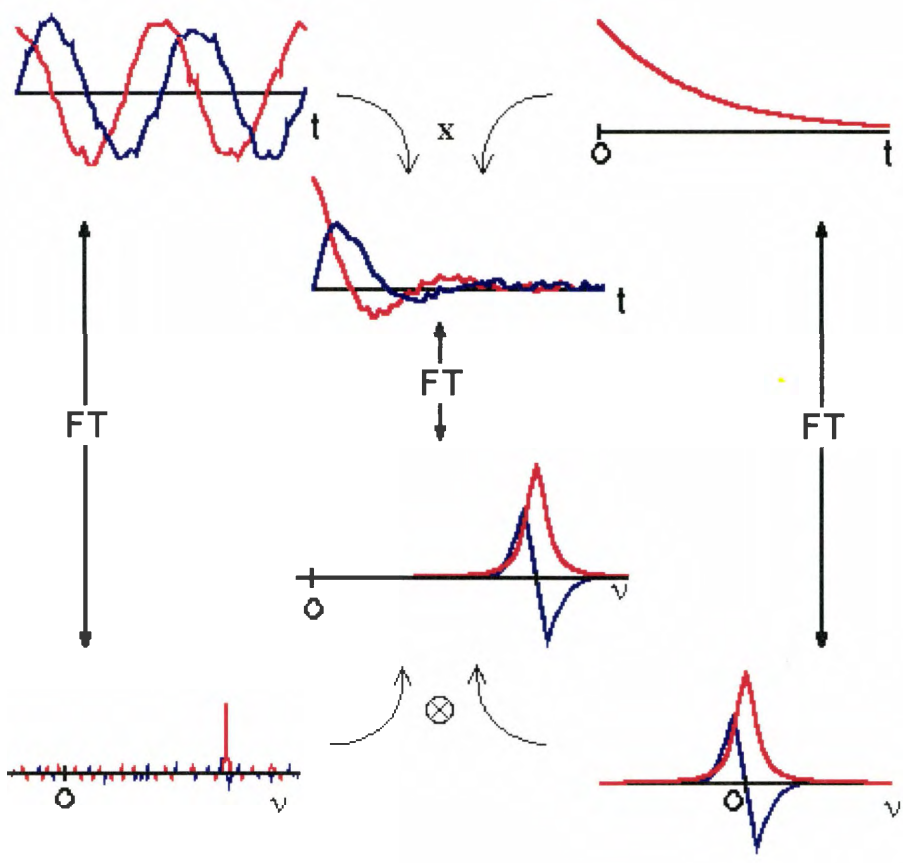


To the magnetic resonance scientist, the most important theorem concerning Fourier transforms is the convolution theorem. The convolution theorem says that the Fourier Transform of a convolution of two functions is proportional to the products of the individual Fourier transforms, and vice versa.

If $f(\omega) = FT(f(t))$ and $h(\omega) = FT(h(t))$, then,

$$f(\omega)\, g(\omega) \;=\; FT(g(t) \otimes f(t)) \;\; and \;\; f(\omega) \otimes g(\omega) \;=\; FT(g(t)\, f(t))$$

Another application of the convolution theorem is in noise reduction. With the convolution theorem it can be seen that the convolution of an NMR spectrum with a Lorentzian function is the same as multiplying the time domain signal by an exponentially decaying function.



(Source: http://www.cis.rit.edu/htbooks/nmr)

In a nuclear magnetic resonance spectrometer, the computer does not see a continuous FID, but rather an FID which is sampled at a constant interval. Each data point making up the FID will have discrete amplitude and time values. Therefore, the computer needs to take the Fourier Transform of a series of delta functions which vary in intensity.

The wrap around problem or artifact in a nuclear magnetic resonance spectrum is the appearance of one side of the spectrum on the opposite side. In terms of a one dimensional frequency domain spectrum, wrap around is the occurrence of a low frequency peak which occurs on the high frequency side of the spectrum.

The convolution theorem can explain why this problem results from sampling the transverse magnetization at too slow a rate.

The two-dimensional Fourier transform (2-DFT) is a Fourier Transform performed on a two dimensional array of data. The 2-DFT is required to perform state-of-the-art MRI.

2.2 NMR Structure determination and refinement

Proteins are biopolymers made up of twenty different amino acids, each having an acid group, an amino group, and a side-chain. The order of the amino acids and the properties of their side chains in a protein determine a three-dimensional structure, which specifies the function of the protein.

We study a problem related to the NMR approach of protein structure determination. More specifically, we want to proceed from a set of inter-atomic distances to a three dimensional molecular structure, which corresponds to the minimum energy structure of the protein. The inter-atomic distances are estimated from NMR experiments.

So the problem formulation can be viewed as, trying to find the Euclidean co-ordinates of the atoms so that the distances between the pairs of atoms are within the corresponding constraints (bounds) determined or set by us.

We can attain this goal by applying approaches from distance geometry and optimization.

There are essentially three steps in the NMR structure determination:

1. Bound Smoothing
2. Embedding
3. Optimization

The first two are essentially what is know as the distance geometry problem, from which we obtain an embedded structure. The final step is that which incorporates stochastic simulated annealing methods to reduce/remove any distortions in the obtained structure. The following theorem is of great importance in this field of application.

Theorem (Saxe):

When all the exact distances are given, the distance geometry problem can be solved in polynomial time.

If the set of distances is sparse, then the N-dimensional distance geometry problem is NP-hard for all $N \geq 1$,i.e., it is not possible to find a polynomial time algorithm to solve the problem.

So the distance geometry problem, when all the exact distances are given, can be solved in polynomial time ($O(n^3)$ floating point operations).

In NMR experiments, though, the set of distances is usually always sparse, as some distances cannot be estimated from NMR experiments, and there are often experimental errors as well. Because of the experimental errors, thus, the distances are specified as pairs of bounds (ranges).

Then, in the distance geometry problem, we "tighten the intervals" using a set of inequalities the distances have to satisfy (this is called "bound smoothing"). Then distances are chosen at randomly within the bounds, and the so-called metric matrix is generated. "Embedding" then converts this matrix to three-dimensional co-ordinates.

Bound Smoothing

Bound smoothing can be used to estimate missing data or correct inconsistencies in given distance bounds. The method is based on geometric rules such as triangular inequalities, as well as physical principles such as the Van der Waal's spheres between the pairs of atoms.

One of the geometric rules used is the following.

Suppose that the lower and upper bounds for two of the distances in between points i, j and k are given. Let the bounds be denoted as l_{ij} , u_{ij} , l_{jk}, and u_{jk}. Then the lower and upper bounds for the distance between points i and k must agree with the following rules,

$$l_{ik} = \max \{ l_{ik}, l_{ij} - u_{jk} , l_{jk} - u_{ij} \}$$

$$u_{ik} = \min \{ u_{ik}, u_{ij} + u_{jk} \}$$

Other rules can also be derived similarly for the distance bounds within more than three points.

Embedding

Bound smoothing only reduces the possible intervals for inter-atomic distances from the original bounds. However, the embedding algorithm demands a specific distance for every single atom pair in a molecule. These distances are chosen randomly within the interval to generate a trial distance matrix, $D = [d_{ij}]$ where d_{ij} is the distance between atoms i and j.

From this we generate the metric matrix, M, which is the matrix of all scalar products of position vectors of the atoms when the geometric center is placed at the origin, i.e.,

$M = [M_{ij}]$ where $M_{ij} = ½ (d_{i0}^2 + d_{j0}^2 + d_{ij}^2)$, d_{i0} is the distance of atom i from the origin.

This is reduced to the calculation of eigenvalues of M using Singular Value Decomposition (SVD), which can be done in $O(n^3)$ floating point operations.

$$(M_{ij})\,\vec{e} = \lambda\,\vec{e}$$
$$\sqrt{\lambda_1}\ \vec{e}_1 = (x_1, x_2, \ldots, x_n)$$
$$\sqrt{\lambda_2}\ \vec{e}_2 = (y_1, y_2, \ldots, y_n)$$
$$\sqrt{\lambda_3}\ \vec{e}_3 = (z_1, z_2, \ldots, z_n)$$

If the solution exists, then the rank(M) = 3, then there are only three non-zero eigen values of M and the above gives the x-,y- and z- co-ordinates of the atoms. In practice though, the rank(M) is not 3. There will be more than three positive eigen values. In such a case, the eigen value expansion is truncated after the largest(first) three eigen values λ_1, λ_2 and λ_3, and the corresponding eigen-vectors are e_1, e_2 and e_3. This corresponds to a projection of a higher-dimension object into three-dimension space.

Optimization

Refinement of the embedded structure is always necessary to remove distortions in the structure. During the projection, bond lengths are distorted. The embedded structure obtained from distance geometry as described above, is then used as the initial structure in the simulated annealing procedure to obtain a refined structure.

Simulated Annealing

Simulated annealing is a special case of either Molecular Dynamics, Langevin Dynamics, or Monte Carlo simulation. As its name implies, the Simulated Annealing (SA) exploits an analogy between the way in which a metal cools and freezes into a minimum energy crystalline structure (the annealing process) and the search for a minimum in a more general system. It is an alternative stochastic optimization technique that mimics the physical process by which a crystal is grown from a melt. Physical systems may be coaxed into a minimum energy conformation (e.g., crystal) by a slow annealing process. Often, the system is first heated and then cooled. Thus, the system is given the opportunity to surmount energetic barriers in a search for conformations with energies lower than the local-minimum energy found by energy minimization. If the reduction of the temperature is slow enough, the system is able to pass out of local energy minima, and arrives at the global minimum configuration. This principle may be applied to solve optimization problems. This improved equilibration can lead to more realistic simulations of dynamics at low temperature. Of course, annealing is more expensive than energy minimization. Simulated annealing is often applied to potentials, V(R), that include unphysical energy terms, as when annealing structures to reduce crystallographic R factors.

SA's major advantage over other methods is an ability to avoid becoming trapped at local minima. The algorithm employs a random search which not only accepts changes that decrease objective function **f**, but also some changes that increase it. The latter are accepted with a probability

$$p = e^{\frac{-\delta f}{T}}$$

where δf is the increase in f and T is a control parameter, which by analogy with the original application is known as the system `temperature' irrespective of the objective function involved. The simulated annealing algorithm is based on that of Metropolis et al., which was originally proposed as a means of finding the equilibrium configuration of a collection of atoms at a given temperature.

The implementation of the SA algorithm is remarkably easy. The flowchart below shows its basic structure. The following elements must be provided:

i. a representation of possible solutions

ii. a generator of random changes in solutions

iii. a means of evaluating the problem functions

iv. an *annealing schedule* - an initial temperature and rules for lowering it as the search progresses.

The potential energy function subject to minimization is given by:

$$E = E_{nb} + E_{bl} + E_{ba} + E_{ta} + E_{imp} + \sum w_a \left(|E_{obs}| - k|E_{calc}|\right)^2$$

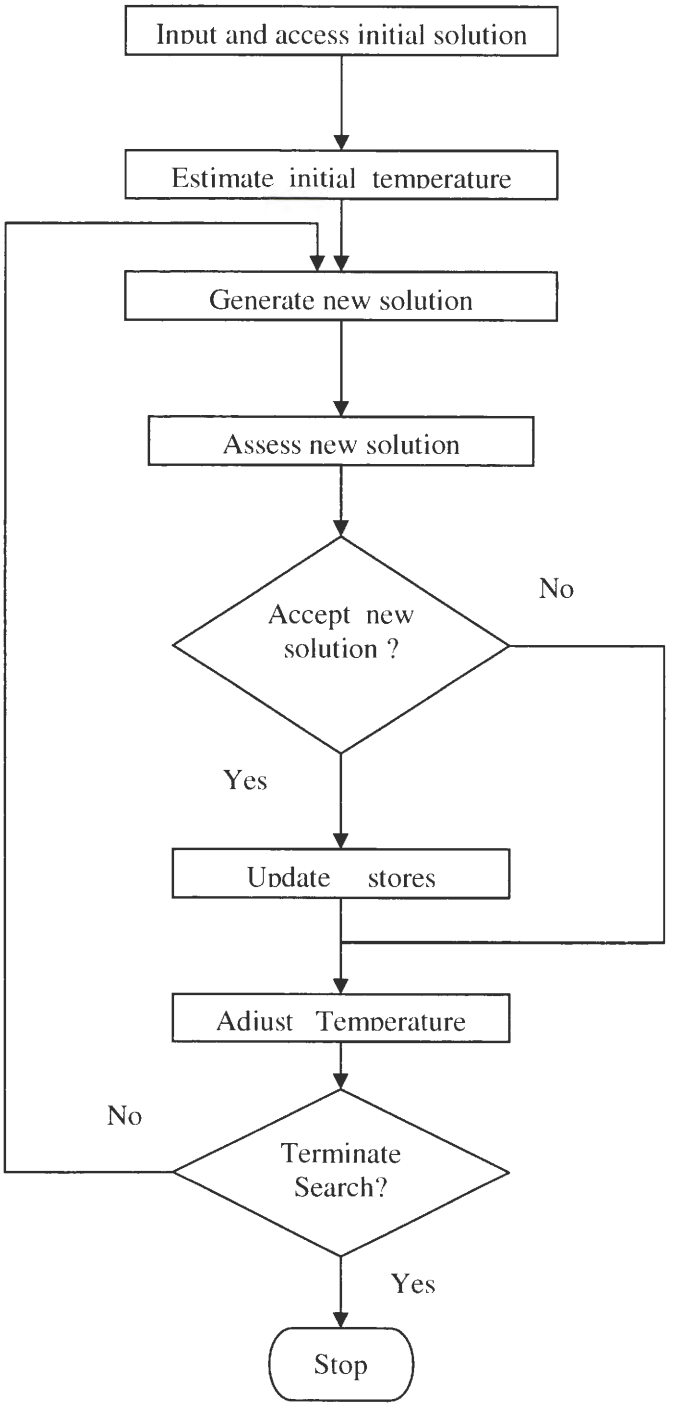
Where $E_{nb} = \sum_{ij\text{-}nonbonded} \varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]$, $E_{imp} = \sum_{improper} k_\theta (\theta - \theta_0)^2$

$E_{bl} = \sum_{ij-bonded} c_{ij} (r_{ij} - r_{ij}^\circ)^2$, $E_{ba} = \sum_{ijk\text{-}bonded} c_{ijk} (\cos\theta_{ijk} - \cos\theta_{ijk}^\circ)^2$

$E_{ta} = \sum_{ijkl\text{-angle}} c_{ijkl}[1 + \cos(n\theta_{ijkl} - \theta_{ijkl}^\circ)], \qquad n = 2, 3$

$r_{ij}, \theta_{ijk}, \theta_{ijkl}$ -- *variables*

$\varepsilon_{ij}, \sigma_{ij}, c_{ij}, r_{ij}^\circ, c_{ijk}, \theta_{ijk}^\circ, c_{ijkl}, \theta_{ijkl}^\circ$ -- *parameters*



The structure of the Simulated Annealing Algorithm

Source: http://csep1.phy.ornl.gov/CSEP/GIFFIGS/MOF217.GIF

Simulated Annealing Algorithm:

```
input initial x0;
y0 = f (x0);
set x = x0; y = y0;
for T = T0, T1, ..., Tm (decreasing)
    for k = 1, ..., n
        x1 = perturb (x0);
        y1 = f (x1);
        dy = y1 – y0;
        e = exp (- dy / T);
        if (rand < e)
          x0 = x1; y0 = y1;
        end
        update x, y;
    end
end
```

SA generates a trajectory through the search space by making incremental changes to a single set of problem parameters, i.e., the torsion angles. At the start of the run these parameters are initialized at random.

Here is a MATLAB script file illustrating the simple version of the annealing algorithm:

```
function [x, y] = annealing2simple (f, x0)
y0 = feval (f, x0);
x = x0; y = y0; alpha = 0.9; T = 2; s = 2;
```

```
for k = 1 : 20
  T = alpha * T;
  accept = 0;
  for l = 1 : 100
    x1 = x0 + (0.5 – rand (size (x0))) * s;
    y1 = feval (f, x1);
    dy = y1 – y0;
    if rand < exp (- dy / T)
      x0 = x1; y0 = y1; accept = accept + 1;
      if y0 < y, x = x0; y = y0; end
    end
  end
  if accept < 25, s = s / 2; end
  if accept > 75, s = 2 * s; end
  disp ([’# accept steps: ’, num2str(accept)]);
  disp ([’current energy: ’, num2str(y)]);
end
```

The potential function defined above is the Lennard-Jones potential function:

```
function [f, g] = ljfun (x)
n = max (size (x)) / 3;
f = 0; g = zeros(3*n, 1);
for i = 1 : n
  for j = i+1 : n
    i1 = 3*(i-1) + 1; i2 = i1 + 1; i3 = i2 + 1;
    j1 = 3*(j-1) + 1; j2 = j1 + 1; j3 = j2 + 1;
```

```
    r1 = x(i1)-x(j1); r2 = x(i2)-x(j2); r3 = x(i3)-x(j3);
    rr = r1 * r1 + r2 * r2 + r3 * r3;
    r = sqrt (rr);
    r6 = 1 / rr / rr / rr;
    f = f + (r6 - 2) * r6;
    dr = - 12 * (r6 - 1) * r6 / r;
    g(i1) = g(i1) + dr * r1 / r;
    g(i2) = g(i2) + dr * r2 / r;
    g(i3) = g(i3) + dr * r3 / r;
    g(j1) = g(j1) - dr * r1 / r;
    g(j2) = g(j2) - dr * r2 / r;
    g(j3) = g(j3) - dr * r3 / r;
  end
end
```

In order to run the algorithm, we need to minimize the potential:

```
rand ('state', 0);
x0 = rand (36, 1);
[x1, y1] = annealing2simple ('ljfun', x0);

x1
y1
```

Note, that the above script file is for a single run of the algorithm. Multiple runs can be adopted to improve the results.

One can then amend the efficiency of the algorithm by decreasing the step size, increasing the number of cooling steps, resetting the starting states etc.

The NMR refinement is an iterative process of improving agreement between the molecular model and NMR data. Protein structure determination by NMR ends with building a model of the protein that fits distance restraints from multi-dimensional NMR spectra. This is no trivial task. One general procedure entails starting from a model of the protein having the known sequence of residues, and having standard bond lengths and angles but random conformational angles. This starting structure will, of course, be inconsistent with most of the distance and conformational restraints derived from NMR. The amount of inconsistency can be expressed as a numerical parameter that should decline in value as the model improves, in somewhat the same fashion as the R-factor decreases as a crystallographic model's agreement with diffraction data improves during crystallographic refinement.

Starting from a random conformation, simulated annealing or some form of molecular dynamics is used to fold the model under the influence of simulated forces that maintain correct bond lengths and angles, provide weak versions of van der Waals repulsions, and draw the model toward allowed conformations, as well as toward satisfying the restraints derived from NMR. Electrostatic interactions and hydrogen-bonding are usually not simulated, in order to give larger weight to restraints based on experimental data; after all, we want to discover these interactions in the end, not build them into the model before the data have had their say.

The resulting model is examined for serious van der Waals collisions, and for large deviations from even one distance or conformational restraint. Models that suffer from one or more such problems are judged not to have converged to a satisfactory final conformation. They are discarded. The entire simulated folding process is carried out repeatedly, each time from a different random starting conformation, until a number of models (an ensemble) are found that are chemically realistic and consistent with all NMR-derived restraints. When the group of models appears to contain the full range of structures that satisfy all restraints, this phase of structure determination is complete. Finally, a single model, structurally averaged/energy minimized model, is derived from the ensemble.

Root mean square deviations (RMSD) from average ensemble coordinate positions (NMR)

Root mean square deviation (RMSD), is a measure to determine how much the position of each atom in a model varies throughout the ensemble. The RMSD for an atom is the square root of the sum of squares of distances between that atom in all models in an NMR ensemble and the average position for that atom in the ensemble. The best quality models exhibit main-chain deviations no greater than 0.4 Å, with side-chain values below 1.0 Å. An averaged model can be colored according to this criterion. For emphasis: such coloring DOES NOT reflect the distances of averaged-model atoms from the average, but instead the amount of variation in atom positions in the ensemble.

Structurally averaged/energy minimized model (NMR)

A single macromolecular model derived from an ensemble of NMR models by averaging atom positions and minimizing the energy (Fig 2.2).

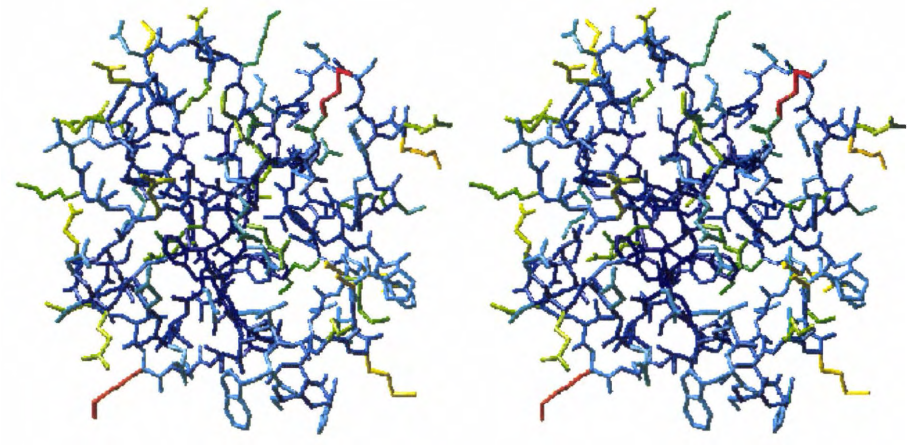


Figure 2.2: Structurally averaged and energy minimized model of human thioredoxin (PDBID: 3TRX)

Source: http://www.usm.maine.edu/~rhodes/ModQual/#Ramachandran%20diagram

In Figure 2.2, atoms are colored by atomic RMS deviation of individual models about the mean atomic positions. Specifically, in red areas, the variation among the 33 models is greatest, and in blue areas, the variation among models is the smallest.

A general procedure is to compute the average position for each atom in the model and to build a model of all atoms in their average position. This model may be unrealistic in many respects. For example, bond lengths and angles involving atoms in their averaged positions may not be the same as standard values. This averaged model is then subjected to restrained energy minimization, which in essence brings bond lengths and angles to standard values, minimizes van der Waals repulsions, and maximizes non-covalent interactions, with minimal movement away from the averaged atomic coordinates.

In summary, one should note that *models are not molecules observed.* No matter how they are obtained, before we ask what they tell us, we must ask how well macromolecular models fit with other things we already know. A model is like any scientific theory: it is useful only to the extent that it supports predictions that we can test by experiment. Our initial confidence in it is justified only to the extent that it fits what we already know. Our confidence can grow only if its predictions are verified.

2.3 Protein conformation framework

Most proteins assume a distinct three-dimensional structure in vivo and vitro, and this native structure is necessary for function. A protein's structure is determined by its amino acid sequence and the surrounding environment.

There are various levels of structure(Figure 2.3):

-Primary structure: refers to the amino acid sequence

-Secondary structure: refers to the fold of the peptide chain, i.e. α-helices and β-strands and turns

-Tertiary structure: describes the 3-dimensional structure and is determined by the packing of the secondary structural elements and the amino acid side chains. It describes the overall shape of the protein, caused by interactions between the various local structures.

Some proteins have a quaternary level of organization, which is defined by the interactions between the tertiary structures of two or more protein subunits (Figure 2.3).

As a protein or peptide unfolds, interactions are disrupted, and both the secondary and tertiary structures can be lost. In addition, the folded and unfolded states of proteins are in equilibrium. Even under conditions that favor the folded state, a protein is not locked into a single conformation. The amino acids are free to interact and move according to the forces placed on them by neighboring atoms and the solvent, producing many conformations that may differ only slightly (conformers).

Protein motion occurs on a wide spectrum with respect to time. Bond vibrations and rotations occur on a femto-second timescale (10^{-15} s), whereas amino acid side chains move at a pico-second to nano-second timescale (10^{-12} – 10^{-9} s). Small segments of peptide bone can reposition themselves in a matter of nanoseconds to microseconds (10^{-9}-10^{-6} s).The arrangement of surface amino acids can be affected by such dynamic behavior. Such changes in the surface of a protein can alter potential binding sites for other molecules. Complete folding of a protein can occur in microseconds to milliseconds or hours, depending on the sequent and solvent environment. Furthermore, proteins and peptides can adopt different folded structures under different conditions.

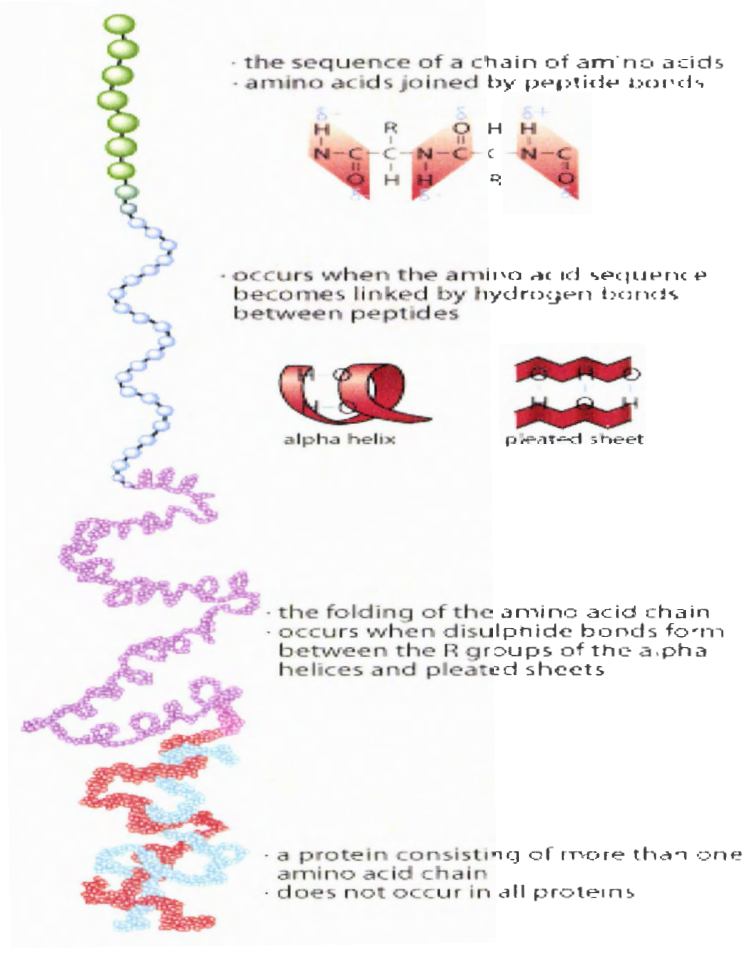


Figure 2.3. Different levels of structural organization of proteins.

Source: http://www.bioteach.ubc.ca/Biomedicine/Prions/

Dihedral Angels

Polypeptides can have a wide variety of conformations, i.e., 3-D structures differing only in the rotational orientations about covalent bonds. This type of rotational flexibility is characterized by a dihedral angle, which measures the relative orientation of four linked atoms in a molecule, i-j-k-l. A dihedral angle for a four-atom sequence that is not necessarily covalently bonded can also be used for special terms in a potential energy function.

The dihedral angle τ_{ijkl} defined for a sequence of linked atoms i-j-k-l is the angle between the normal to the plan of atoms i-j-k and the normal to the plane of atoms j-k-l. Its sign is determined by the triple product (a*b).c, where a, b, and c are the inter-atomic distance vectors for atoms i→ j , j→k , and k→l, respectively.

Strictly defined, the related torsion angle τ' is the angle between the two planes defined by i-j-k and j-k-l. Thus $\tau + \tau' = 180^{\circ}$.

When the dihedral angle is 0°, the four atoms i-j-k-l are coplanar, and i and l coincide in their projections onto the normal plane to the j-k bond (corresponding to the cys orientation). When it is 180°, the atoms are coplanar, but i and l lie opposite on another in the projection onto the plane normal to the j-k bond (corresponding to the trans orientation).

While the peptide is relatively rigid, there is a great deal of flexibility about each of the single bonds along the backbone, N-C_α and C_α-C. The two torsion angles φ and ψ dihedral angles are used to define rotations about the bond between the Nitrogen and C_α of the mainchain and between C_α and the carbonyl carbon, respectively.

The dihedral angle, ω, defines the rotation about the peptide bond, namely for the atomic sequence $C_{1\alpha}$-{C-N}- $C_{2\alpha}$, where C_1 and C_2 are the α-carbons of two adjacent amino acids.

Dihedral angles used to define sidechain rotations are denoted by χ.

Rotameric structures of amino acids are those that have the same φ and ψ angles but differ in sidechain conformations (different value of χ).

Ramachandran diagram

A Ramachandran plot/diagram is a plot showing the main-chain conformational angles in a polypeptide. This diagram is used to find problems in models during structure refinement. The conformational angles plotted are φ, the torsion angle of the N-C_α bond, defined by the atoms C-N-C_α-C (C is the carbonyl carbon); and ψ, the torsion angle of the C_α -C bond, defined by the atoms N- C_α -C-N. In Figure 2.4, $\varphi = \psi = 180^{\circ}$ (convergent stereo).

The feasible combinations of the φ and ψ angles are limited due to steric hindrance. That is, only certain combinations are typically observed, with some dependence on residue size and shape. Glycine is

unique in its flexibility – it is therefore a good agent for turns in polypeptides and proteins – but other residues exhibit a highly limited range of sterically-permissible φ and ψ combinations In fact, only roughly $1/10^{th}$ the area of {φ, ψ} space is generally observed for polypeptides and proteins. Among the first to note this limitation were G.N.Ramachandran and co-workers, after whom the Ramachandran plots are named. Around the same time, Johj Schellman and co-workers were working independently along the same lines of mapping the energetically favorable and excluded regions for protein conformations.

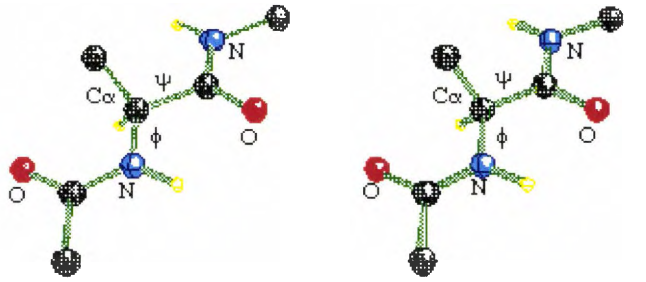


Figure 2.4. Main-chain conformational angles in a polypeptide

Source: http://www.usm.maine.edu/~rhodes/ModQual/#Ramachandran%20diagram

The allowed pairs of values are depicted on a Ramachandran diagram as irregular polygons that enclose backbone conformational angles that do not give steric repulsion (inner polygons) or give only modest repulsion (outer polygons).

Every point (phi, psi) on the diagram represents the conformational angles phi and psi on either side of the C_α of one residue. Each residue in the protein is represented with a dot or other mark on the plot.

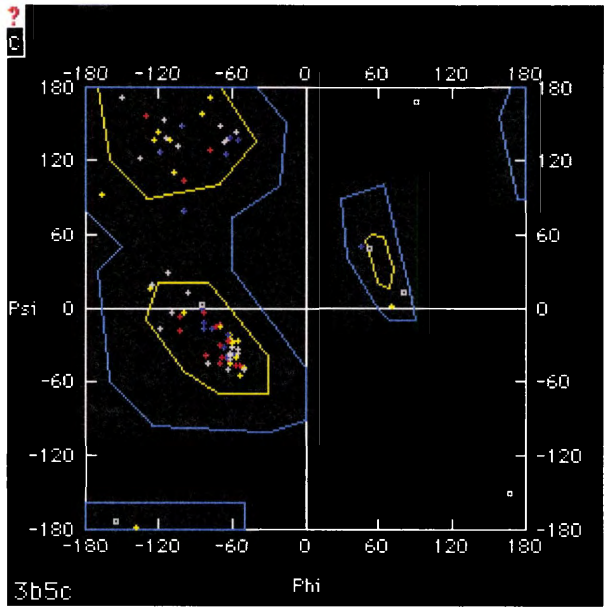


Figure 2.5 :Ramachandran diagram for cytochrome b5 (PDB 3b5c). Small squares represent glycine residues; small crosses represent all others. Residues are colored by type: blue = positive, red = negative, yellow = polar, gray = nonpolar. Note that, in this very well-refined model, only glycines lie outside of allowed regions (blue polygons).

Source: http://www.usm.maine.edu/~rhodes/ModQual/#Ramachandran%20diagram

These diagrams/plots in the {φ, ψ} space are used to describe this {φ, ψ} flexibility (or rather inflexibility) in polypeptides and proteins. Below is another figure showing what a Ramachandran plot.

During the final stages of map fitting and crystallographic refinement, Ramachandran diagrams are useful to find conformationally unrealistic regions of a protein model. Structure publications often include the diagram, with an explanation of any residues that lie in "forbidden" areas or unfavorable regions, which correspond to high-energy conformations. Due to the lack a side chain, Glycines, usually account for most of the residues that lie outside allowed regions. If non-glycine residues exhibit forbidden conformational angles, there should be some explanation, such as structural constraints that overcome the energetic cost of an unusual backbone conformation.

Often, Ramachandran diagrams are presented by plotting the backbone dihedral angles of all non-terminal residues in a protein for a large group of known protein structures. This superimposed view, averaged

over many residues, approximates protein conformation tendencies. The favorable regions correspond to common secondary-structure elements such as helices and sheets, with finer motifs also noted.

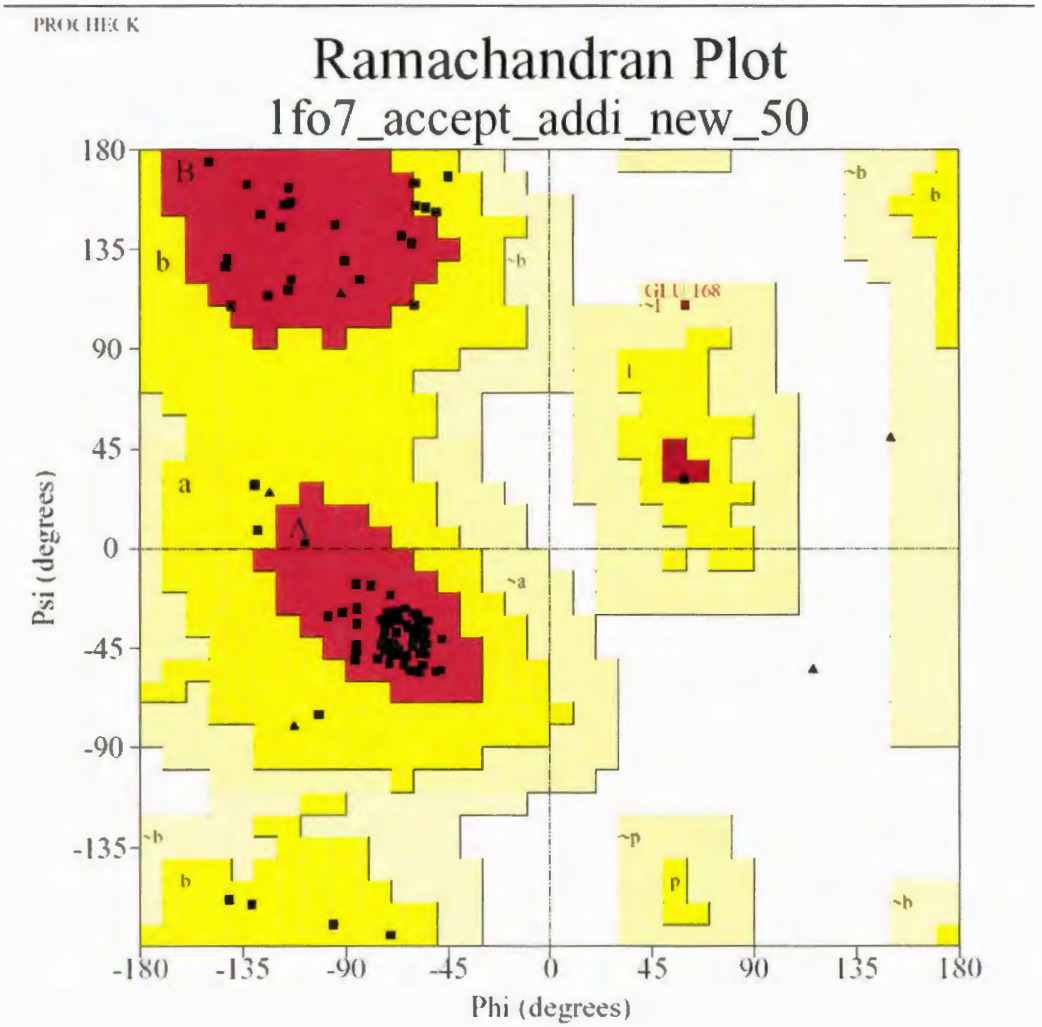


Figure 2.6: Ramachandran plot for 1FO7 average minimized plot we obtained after adding additionally database derived constraints.

2.4 Human Prion Protein

Transmissible Spongiform Encephalopathies

The transmissible spongiform encephalopathies (TSEs) or Prion diseases are a group of neuro-degenerative disorders which include Creutzfeldt-Jakob Disease (CJD), Gerstmann-Straussler-Scheinker disease (GSS) and fatal familial insomnia in humans, and mad cow disease in bovines, scrapie in sheep, goats and mufflons, chronic wasting disease of deer and elk, and feline spongiform encephalopathy in domestic cats.

CJD was first described in the 1920's and is the most common human spongiform encephalopathy. The prion diseases can be:

-sporadic (no known cause or source)

-iatrogenic sources (acquired from environmental sources)

-familial (due to mutations in a gene).

GSS and fatal familial insomnia in humans, are two exclusively familial disorders, and are rarer.

All the prion disorders are experimentally transmissible. At autopsy, microscopic analyses of brain sections show widespread spongiform changes accompanied by a reactive astrocytic gliosis and neuronal loss. Immunochemical staining reveals the presence of prion protein amyloid deposits or plaques. Structural transition of the normal prion protein to an aberrant form which is proteinase K- resistant, causes prion diseases.

Table 2.1 Illustrates different types of Prion diseases.

Disease	Abbreviation	Affected
Creutzfeldt-Jakob Disease	CJD	Humans
Variant Creutzfeldt-Jakob Disease	vCJD	humans; acquired from cattle with BSE
Bovine Spongiform Encephalopathy	BSE	"mad cow disease" in cattle
Kuru		infectious; in humans who practiced cannibalism in Papua New Guinea
Gerstmann-Sträussler-Scheinker disease	GSS	inherited disease of humans
Fatal Familial Insomnia	FFI	inherited disease of humans
Scrapie		infectious disease of sheep and goats
other animal TSEs		Cats, mink, elk, mule deer

Source:http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/P/Prions.html

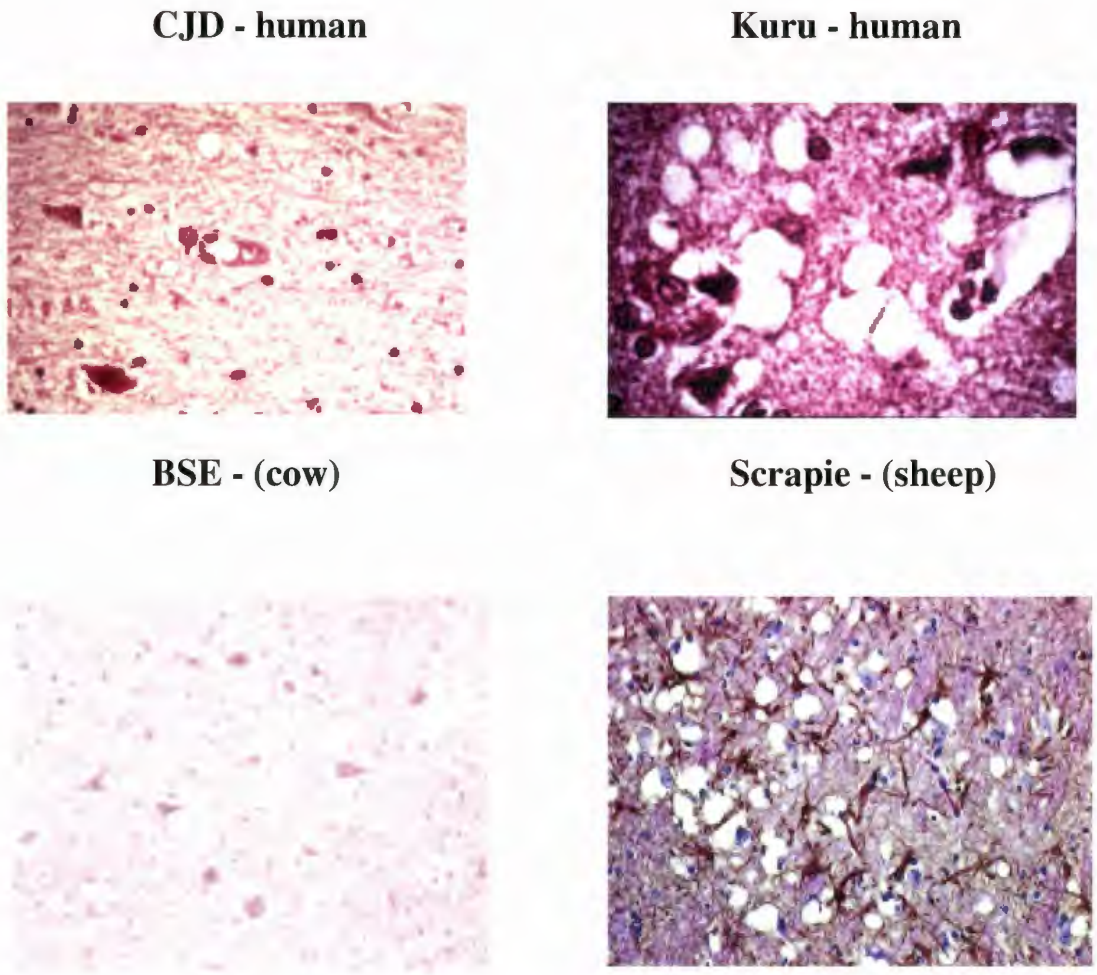


Figure 2.7 Histopathology of Prion Diseases

Source:http://www.cyber-dyne.com/~tom/braingifs.html

What Are Prions?

The term prion comes from "proteinaceous infectious particles".

Prions are proteins found on the plasma membrane (the membrane that surrounds a cell and defines its physical boundary). In mammals, prions are found in the highest concentration in cells of the central nervous system.

Prions, like all proteins, can have up to four levels of structural organization. The normal cellular form of the protein is denoted as PrP^{C} , while the aberrant or mutant form as PrP^{Sc} .The function of PrP^{C} is unknown. PrP^{Sc} is thought to be the causative agent of the set of neurological disorders comprising the prion diseases.

The prion protein (PrP) is encoded by a single exon of a single-copy chromosome gene. In addition to the protein-coding exon, PrP genes in mammals contain one or two 5`-noncodings exons.

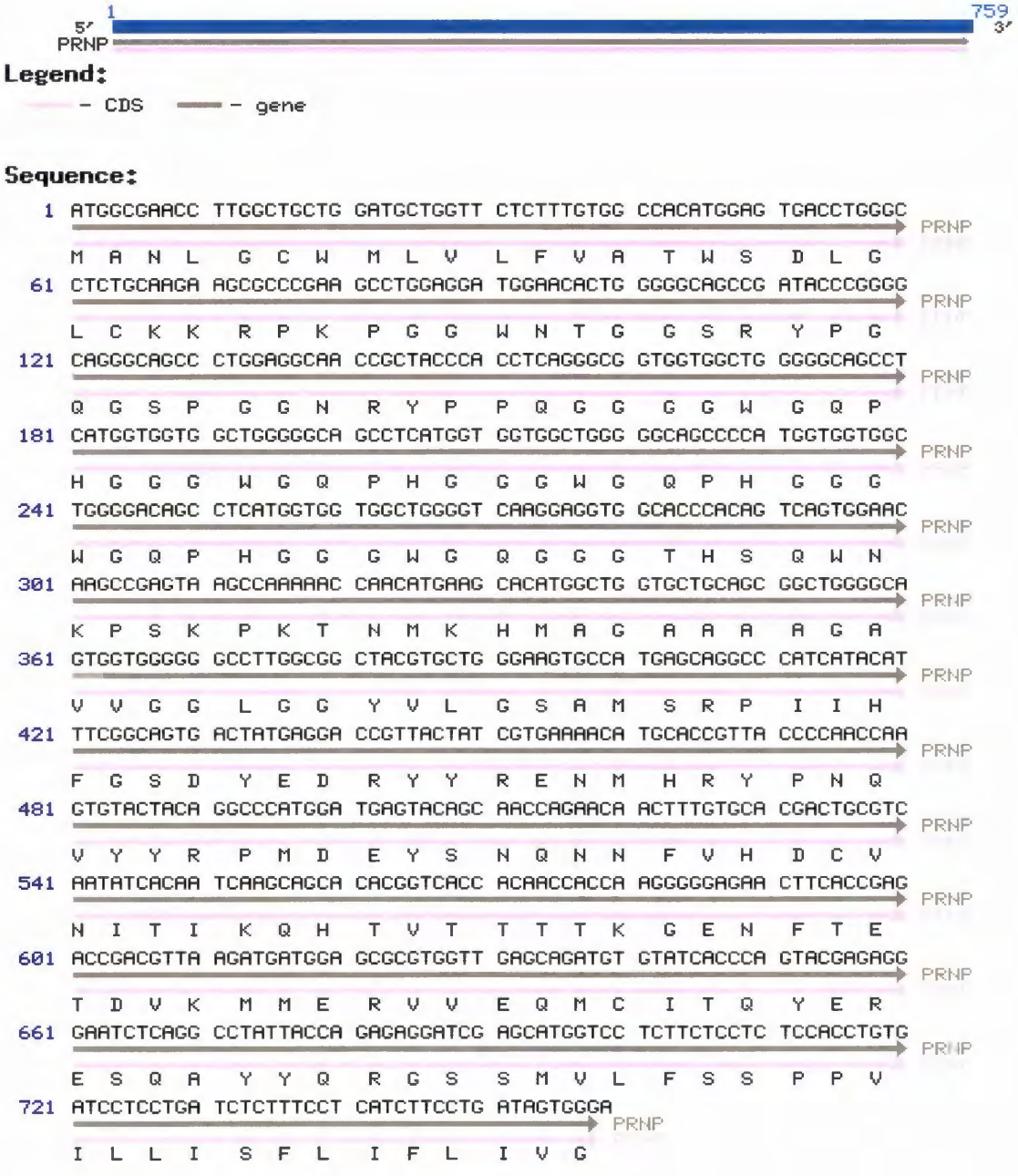


Figure 2.8. The primary structure of Human Prion gene (PRNP)

Some prion diseases have been identified and linked with point mutations in the prion gene (PRNP).

Table 2.2 Mutations and prion diseases

Mutation	**Disease Phenotype**
Octa-repeat insertion of 24, 48, 96, 120, 144, 168, 192, or 216 base pairs between codons 51 and 91	CJD, GSS, or atypical dementias
P102L (Pro Leu)	GSS: classical ataxic form
P105L (Pro Leu)	GSS: spastic paraparetic variant
A117V (Ala Val)	GSS: pseudobulbar variant
G131V (Gly Val)	GSS: classical ataxic form
Y145* (Tyr Stop)	Alzheimer-like dementia
D178N (Asp Asn)	CJD (129V on mutant allele)
D178N (Asp Asn)	FFI (129M on mutant allele)
V180I (Val Ile)	CJD
T183A (Thr Ala)	Alzheimer-like dementia

H187R (His Arg)	GSS: classical ataxic form
F198S (Phe Ser)	GSS with neurofibrillary tangles
E200K (Glu Lys)	CJD
D202N (Asp Asn)	GSS with neurofibrillary tangles
V203I (Val Ile)	CJD
R208H (Arg His)	CJD
V210I (Val Ile)	CJD
E211Q (Glu Gln)	CJD
Q212P (Gln Pro)	GSS with Lewy bodies
E217R (Glu Arg)	GSS with neurofibrillary tangles
M232R (Met Arg)	CJD

Abbreviations: CJD = Creutzfeldt-Jakob disease; GSS = Gerstmann-Straussler-Scheinker syndrome; FFI = fatal familial insomnia.

Source: http://www.cjdinsight.org/familialcjd.html

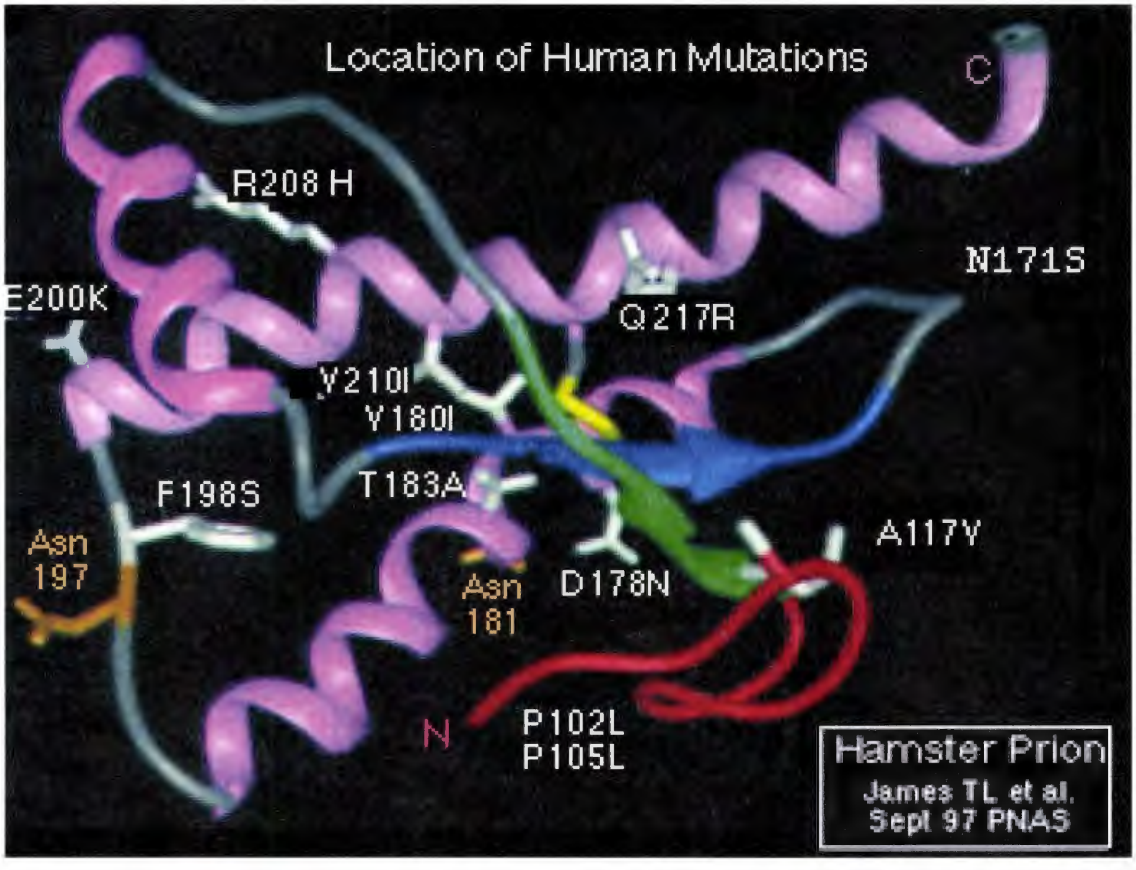


Figure 2.9 Illustrates the loci on the prion protein resulting from the different point mutations.

Source:http://www-micro.msb.le.ac.uk/3035/prions.html

The 'Protein Only Hypothesis'

Prion diseases are some of the most intriguing diseases affecting the brains of humans and animals. The prevalent controversial hypothesis proposes that the infectious agent is a misfolded protein that propagates in the absence of nucleic acid by transmission of its altered folding to the normal host version of the protein. That is, it suggests that prion infectivity is associated with a conformational transition of the normal cellular form of the prion protein (PrP^{C}) to its isoform (PrP^{Sc}) without known life pathogens.

Several other models of prion infectivity have been hypothized, but data is strongly in favor towards the 'protein only' hypothesis, though it should be noted that the final proof- consisting of the generation of infectious prions in vitro – is still missing.

The scrapie agent had been found (by Alper et al) to be extremely resistant to inactivation by UV and ionizing radiation, as shown in Figure 3. This supports the protein-only hypothesis further.

In view of this evidence, and accepting the hypothesis, it is suggested that the transition from PrP^{C} to PrP^{Sc} is what causes prion infectivity.

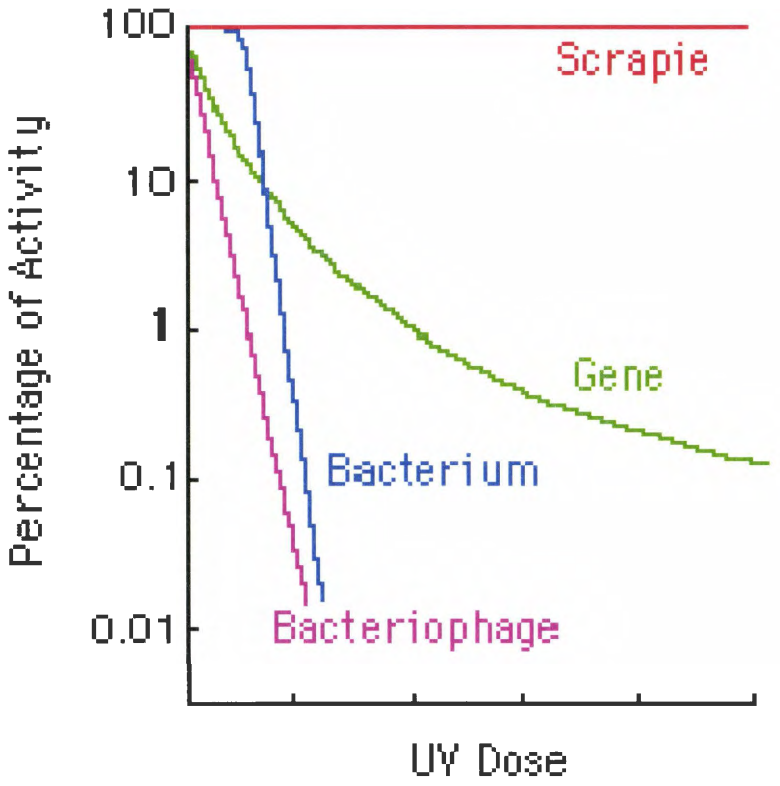


Figure 2.10 Resistance of the scrapie agent to UV Ionization

Source : http://cloud.prohosting.com/lzambeni/prions/prion_structure_function.html

PrP^{C} and PrP^{Sc} are isoforms with identical amino acid sequence. However, the two protein isoforms have profoundly different physicochemical and structural properties:

a. PrP^{C} is highly soluble; PrP^{Sc} exists as an insoluble aggregate

b. PrP^{C} is easily degraded by proteinase K; PrP^{Sc} is resistant to proteinase K digestion.

c. PrP^{C} is ~42% helical with a very low (~3%) beta- sheet content; PrP^{Sc} consists of 30% alpha-helixes and 43% beta-sheets.

PrP^{C} is a secretary cell surface glycoprotein made up of the residues 23-231, which is attached the cell membrane via a glycosyl phosphatidylinositol anchor at its C terminus. It has a single disulfide bridge and two glycosylation sites.

The three-dimensional NMR structure of the recombinant HuPrP (23-231) has revealed that the N-terminal segment (23-120) of the protein is flexible and disordered, while the C-terminal domain (121-231) forms three α-helices and a short pair of β-sheets. Because the segment 90-120 is protected from protease digestion in PrP^{Sc}, it was believed to adopt a well defined conformation in the PrP^{Sc} but not in PrP^{C} and appeared to be essential for the transition from PrP^{C} to PrP^{Sc}.

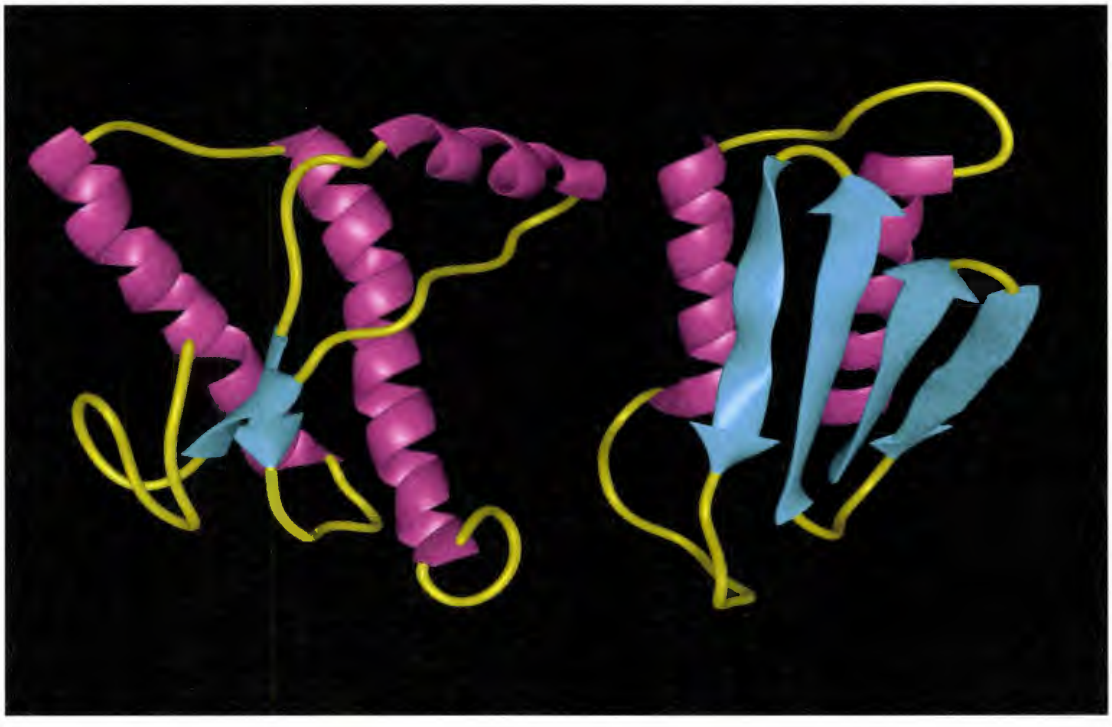


Figure 2.11 Tertiary structure normal prion (left), compared to an aberrant, disease-causing prion (right).

Source: http://cloud.prohosting.com/lzambeni/prions/prion_structure_function.html

To summarize the novel Properties of PrP^C:

-N-linked glycoprotein; normally attached to the cell membrane by a glycophosphatidylinositol (GPI) anchor

-Contains an intra-molecular disulphide bond (Cys 179- Cys 214) that provides structural stability to the C-terminus of the protein.

-Predominantly helical protein

-Unglycosated monomer structure is known by NMR.

Novel properties of the scrapie agent (Prusiner, 1982):

- Stable at 90°C for 30 minutes

- Low molecular weight infectious particles (<50,000 daltons)

- Hydrophobic protein(s) are required for infectivity

- Resistant to ribonucleases and deoxyribonucleases

- Resistant to UV radiation at 254nm

- Resistant to psoralen photoadduct formation

- Resistant to Zn^{2+} catalyzed hydrolysis

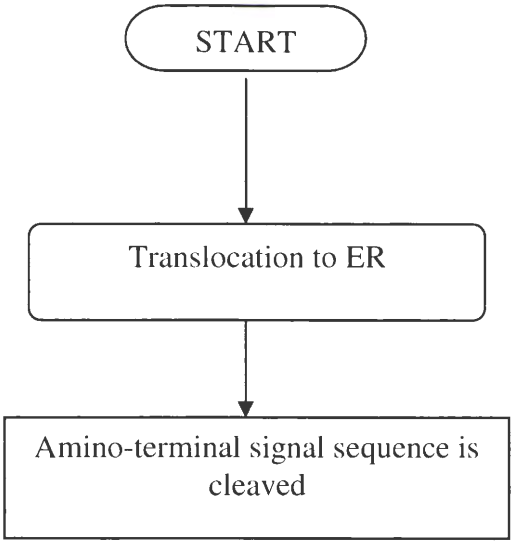
- Resistant to NH_2OH chemical modification

Function of PrP^C

PrP^C is a normal cell surface sialoglycoprotein that is expressed in several tissue types, including neurons and skeletal muscle. The normal physiological function of PrP^C, however, still remains elusive. It has been hypothesized that it is a plasma membrane copper transporter and therefore modulates the activity of certain cuproproteins by affecting the availability of intracellular copper. The conversion into PrP^{Sc} may abolish the protein's ability to bind copper and thus deprive the cell of copper for enzymes that have anti-oxidative activities.

PrP^C binds copper from the extra cellular milieu and transports the metal to the intracellular side of the plasma membrane via the endocytotic pathway. It is likely that the endocytoses of PrP^{Sc} requires the participation of additional membrane associated factors (like Protein X) as PrP^C is anchored to the plasma membrane via a GPI anchor and it does not have transmembrane domains.

The biosynthesis of PrP can be summarized in the flowchart below:



The endoplasmic reticulum (ER) houses the glycosylation and disulphide bon formation of the prion. Proteins that are misfold in the ER go to cytosol where they are degraded by proteosomes . Inhibition of the proteosome causes PrP to accumulate in the cytosol where it can adopt a PrP^{Sc}-like conformation.

How Can Proteins Be Infectious?

Infection of normal cells may occur when an aberrant prion acts as a template for the refolding of a normal prion into a new aberrant prion. It is thought that there is at least one more protein involved: the as-yet-unidentified Protein X. This protein is believed to mediate the folding from a normal into an abnormal prion. When proteins are synthesized inside of a cell there are other special proteins (known as chaperones) that help in this process. Chaperones are proteins that bind to the newly synthesized protein or protein subunit, in order to ensure that the protein is properly folded into its secondary or tertiary structure. It has been hypothesized that Protein X is a type of chaperone.

The concept that a protein can transmit information (i.e. protein structure) is novel. At the time that this idea was proposed, it was truly radical; it conflicted with many years of evidence showing that only DNA was capable of transmitting such information.

This does not mean that prion diseases cannot be genetic. If an individual happens to carry a gene that codes for a mutant prion, then all of the prions that their cells produce will be the abnormal, disease-causing prions. One such disease is familial CJD.

Factors associated with PrP^{C}→ PrP^{Sc} conversion

- metal ion concentrations
- GPI anchor stability
- extra-cellular molecules (e.g. glycosaminoglycans)
- pH
- N-linked glycosylation (can impact the conformation of protein fragments)
- redox environment
- Protein X

The transition might be induced due to the additive effect of a combination of the above factors. Possibly, these factors are inter-related and affect each other. Factors affecting the stability of the disulphide bridge could have a preventive/promotive effect.

Evidence of Protein X

That PrP^{C} interacts with PrP^{Sc} during the formation of nascent PrP^{Sc} was surmised from transgenic (Tg) mouse studies where mice expressing a Syrian Hamster (SHa) PrP transgene were susceptible to SHa prions. When similar Tg mice were produced expressing human (Hu) PrP, no transmission of the prions were formed. However, mice expressing a chimeric Hu/Mo PrP transgene, denoted as MHu2M, were susceptible to HuPrP. In

addition, it was found by Kaneko et al, that Tg mice expressing HuPrP did become susceptible to the prions when they were crossed with PrP-deficient (Prnp $^{0/0}$) mice.

These data taken together argued that it is likely that a molecule other than PrP is involved in the formation of PrP^{Sc}. This molecule was assumed to be a protein and designated 'Protein X'. Although this protein has not yet been isolated, the sites at which it possibly binds to PrP^{C} has been identified.

Based on the results with the MHu2M transgene and earlier studies showing that the N terminus of PrP is not required for PrP^{Sc} formation, it was surmised (Kaneko et al) that the binding of PrP^{C} to Protein X is likely to occur through specific side chains of amino acids located at the C terminus of PrP^{C}.

Binding sites for Protein X

Using scrapie-infected mouse neuroblastoma (ScN2a) cells transfected with chimeric PrP gene in which the mouse C terminus was replaced by human residues, the binding sites have been identified (Figure 2.12). Substitution of a human residue at position 214 or 218 prevented (Mo) PrP^{C} from being converted into PrP^{Sc}. The side chains of residues 214 and 218 protrude from the same surface of the C-terminal α-helix and form a discontinuous epitope with residues 167 and 171 in an adjacent loop. Substitution of a basic residue at position 167,171, or 218 also prevented PrP^{Sc} formation: at a mechanistic level, these mutant PrPs appear to act as 'dominant negatives' by binding to Protein X an functionally sequestering it from the replication process.

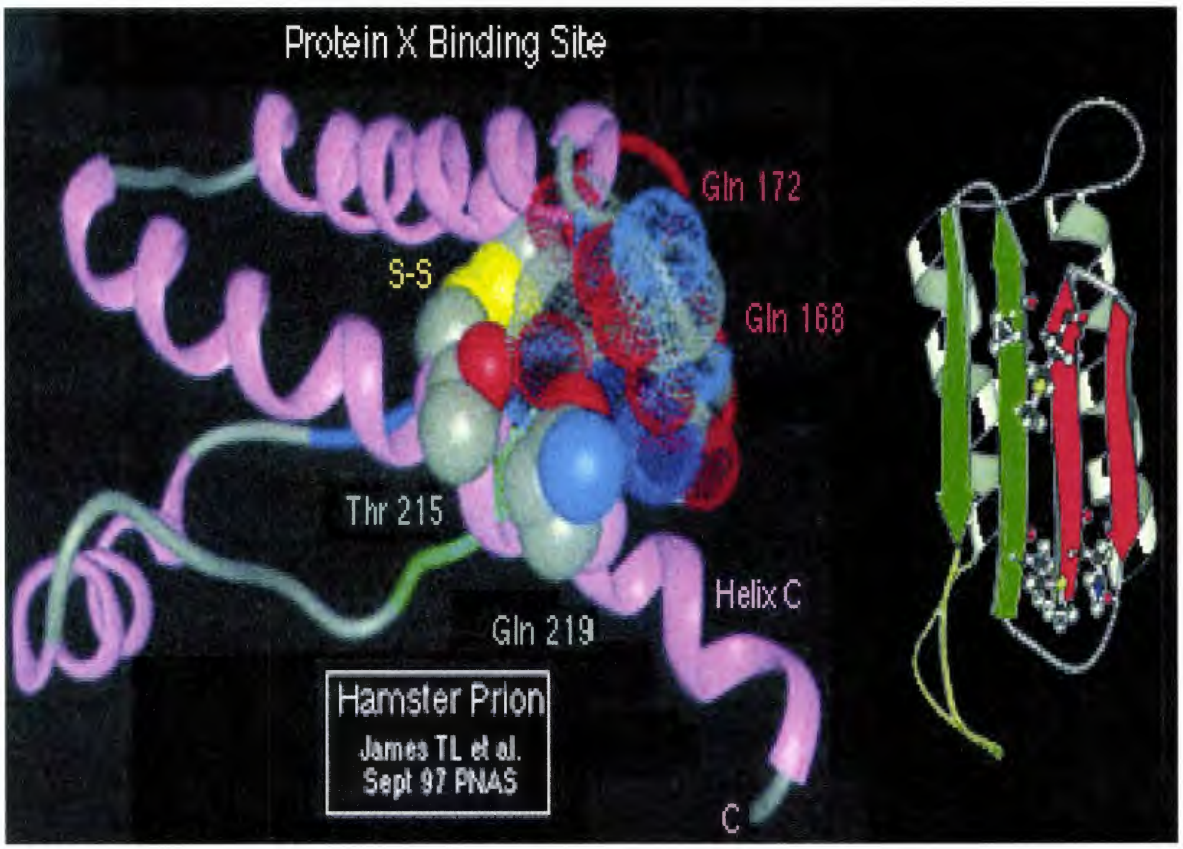


Figure 2.12 Protein X binding site of Human prion protein

Source: http://www1.umn.edu/eoh/hazards/hazardssite/prions/prioncharact.html

Mechanism of Protein X binding in pathogenesis of prion diseases

It is hypothized that PrP^{C} forms a complex with protein X an that PrP^{Sc} then binds to PrP^{C} resulting in a ternary complex. To see this refer to Figure 2.13. The role of protein X in PrP^{Sc} formation and the influence of mutations in PrP^{C} on the prion replication cycle. (*A*) NMR structure of SHa rPrP90-231. The color scheme is as follows: α-helices A (residues 144-157), B (172-193), and C (200-227) in pink; disulfide between Cys-179 and Cys-214 in yellow; hydrophobic cluster composed of residues 113-126 in red; loops in gray; residues 129-134 in green encompassing strand S1 and residues 159-165 in blue encompassing strand S2; the arrows span residues 129-131 and 161-163, as these show a closer resemblance to β-sheet. Structure of protein X binding site of SHa rPrP90-231 illustrating the proximity of the 165-171 loop, where residues Q168 and Q172 are depicted

with a low density van der Waals rendering and helix C residues T215 and Q219 depicted with a high density van der Waals rendering. SHaPrP residues Q168, Q172, T215, and Q219 correspond to MoPrP residues Q167, Q171, T214, and Q218, respectively. Ordering experiments demonstrate that PrP^{C} interacts with protein X prior to the creation of the PrP^{C}/PrP^{Sc} complex. Two cycles are required for PrP^{Sc} formation. The left hand cycle shows protein X binding to PrP^{C} (green) resulting in a heterologous complex that is then competent to interact with PrP^{Sc} (red). Upon conversion of PrP^{C} to nascent PrP^{Sc}, protein X dissociates from the complex owing to its relative lack of affinity for PrP^{Sc}. Protein X is subsequently recycled. The right hand cycle depicts the interaction of PrP^{Sc} with the PrP^{C}/protein X complex and the conversion of PrP^{C} into nascent PrP^{Sc}. With time, the result is an exponential increase in PrP^{Sc} concentration as the template for conversion is recycled.

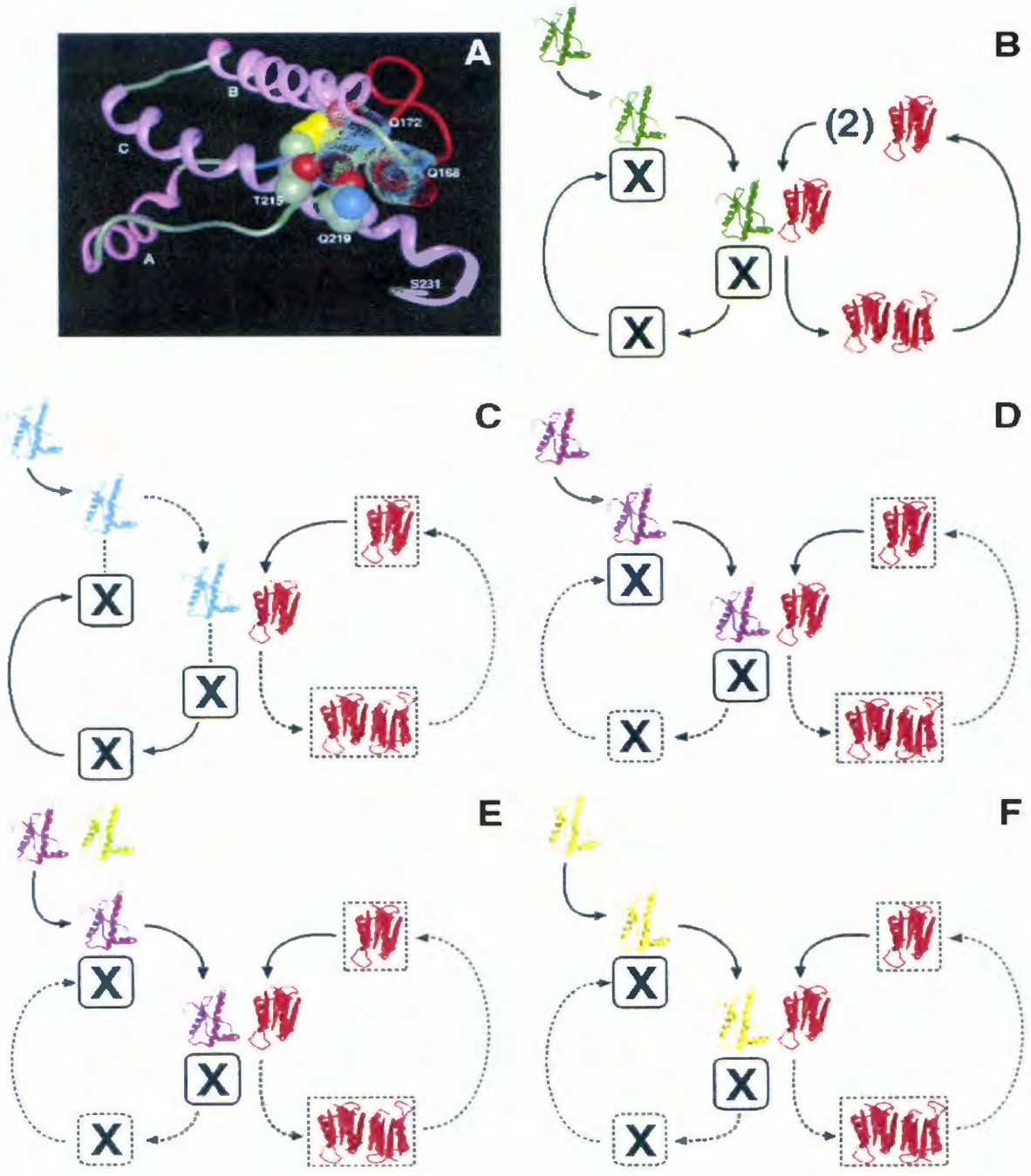


Figure 2.13 Mechanism of protein X binding in pathogenesis of prion diseases:

Source: http://www.pnas.org/cgi/content/full/94/19/10069

(*C*) Type 1 inhibition: mutant PrP^C (blue) containing an amino acid substitution in the PrP^C/protein X interface (e.g., E218 in MoPrP) interacts weakly with protein X. Dotted lines depict the failure of the mutant PrP^C to interact with protein X and the subsequent inability to form the protein X/PrP^C/PrP^{Sc} complex. Under these circumstances PrP^{Sc} formation either does not occur or proceeds slowly. (*D*) Type 2 inhibition: mutant PrP^C (purple) containing an amino acid substitution in the protein X/PrP^C interface (e.g., K218 in MoPrP) forms a very tight complex. PrP^{Sc} is able to bind to this protein X/PrP^C complex but conversion of PrP^C to PrP^{Sc} is prevented owing to the failure of the protein X/PrP^C/PrP^{Sc} complex to release protein X. Dotted lines are shown for the steps in the replication cycle that are blocked. (*E*) Dominant-negative effect of tight binding mutants of PrP^C. Mutant PrP^C [e.g., K218 (purple)] successfully competes with wt PrP^C (green) for binding to protein X. The protein X/PrP^C(K218)/PrP^{Sc} complex is formed but conversion is inhibited as in *D*. (*F*) Type 3 inhibition: PrP^C from a distinct species [e.g., SHa (gold)] is able to bind Mo protein X, but the Mo protein X/SHaPrP^C/MoPrP^{Sc} complex is not competent for conversion. The result is that protein X is sequestered and scrapie prions are not replicated.

How the prion damages its host

The Central Nervous System is the only compartment of the body that undergoes histopathologically and clinically detectable degeneration in prion diseases.

Cellular models of prion disease may be useful to determine how the prion damages the central nervous system, although prions replicate inefficiently in most established cell lines. Many studies have been done using a synthetic peptide from the central region of the PrP^C molecule, which spontaneously assembles into amyloid-like structures.

In vitro, this peptide can elicit reactions that resemble that seen in brain cells during the late stages of prion disease. These include:

- activation of microglial cells
- stimulation of the production of intermediate filaments by astrocytes

- death of neurons

All of the above seem to depend on the presence of the normal prion protein in target cells.

The problem addressed in these studies is: what actually happens in vivo?

Interpretations

-possible that a threshold concentration of PrP^{Sc} is needed for neuro-degeneration and that this level is not reached outside the grafted tissue

-neuronal cytotoxicity of PrP^{Sc} depends on the expression of cellular PrP^{C} by target cells.

-PrP^{C} acts as a receptor of PrP^{Sc} (There is more evidence that this is possible)

-The process of $PrP^{C} \rightarrow PrP^{Sc}$ conversion may be the main deleterious event

Most cellular PrP^{C} is translocated into the lumen of the endoplasmic reticulum(ER) by virtue of its secretory signal peptide. The PrP^{C} is then routed to the cell surface as a GPI-linked membrane-associated protein. But a small portion of PrP^{C} is made a transmembrane form that later leaves the ER. These two forms were named (according to their orientation) - Ctm PrP and Ntm-PrP (carboxy- and amino- transmembrane PrP respectively). It was found (Lingappa et al) that Ctm PrP levels correlate well with neuro-degeneration changes in pathological conditions.This formed two hypotheses:

-Ctm PrP may be a marker of prion-induced neuro-degeneration

-$PrP^{C} \rightarrow PrP^{Sc}$ conversion may trigger the formation of Ctm PrP which may, in turn, be an effector of neuro-toxicity.

Two pathways that prion may reach the central nervous system for neural invasion:

- prions can colonize the immune system as well as lymphocytes

- follicular dendritic cells (which are located in germinal centers and express a considerable amount of PrP^{C})

E200K variant of HuPrP

3-D structures of monomeric and dimeric human prion protein (HuPrP) have been revealed by NMR spectroscopy and X-ray crystallography. However, the NMR-determined structures are not as good in quality as the X-ray structures. This limits the usefulness of the NMR structures. This issue was dealt with using a novel NMR structure refinement approach (discussed in later chapters), which is based on addition of distance constraints derived from a database of high-quality X-ray structures. Because of the importance of understanding the pathogenesis of the transformation of PrP^{C} to PrP^{Sc} that causes diseases like CJD, in this study, we chose the familial CJD-related E200K variant of the Human Prion Protein as a test case to demonstrate the usefulness of this refinement approach.

The structure will be discussed in detail in the methodology when we compare it with the wild-type prion protein also.

CHAPTER 3

MATERIALS AND METHODS

3.1 Methodology

The protein structures determined by Nuclear Magnetic Resonance Spectroscopy (NMR) are not as detailed or as high in quality, as those determined by X-ray crystallography. The reason for this is due to the inadequacy of the distance data available from the NMR experiments. Since the structures are under-determined, this limits the use of the NMR derived structures, which would otherwise give us information regarding the protein's structure in solution. This in turn obstructs the applications in fields like homology modeling and rational drug design.

The distance data from the NMR experiments, can only be obtained for specific atoms (usually hydrogen atoms). It can be estimated as an approximation by defining lower and upper bounds. Thus, instead of determining a single unique structure for a protein, an ensemble of proteins structures is generated. The variation of structures within the ensemble is often considered as a reflection of the flexibility of the protein in solution. But it could be misleading or inaccurate as some variations might be due to structural under-determination.

In order to increase the quality of NMR structures, more distance data has been sought by using various techniques. Some experimental approaches have been developed (such as dipolar coupling), but these are generally computationally expensive.

Theoretical approaches include obtaining additional conformational constraints from databases of known protein structures such as to derive constraints on dihedral angles based on their distributions in known X-ray structures in structural databases (Kuszewski et al). They incorporated the conformational database potential into the target function used for refinement which resulted in significant improvement in various quantitative measures of quality (for instance Ramachandran plots, side-chain angles, overall packing etc.)

In NMR, experimentally derived restraints consist of NOE-derived inter-proton distances (NOE intensities) that may be supplemented by torsion angles, coupling constants and chemical shifts. To be

acceptable, the structure should be in good agreement with the experimental results, while at the same time showing very small deviations from idealized covalent geometry and good non-bonded contacts (interactions between atoms which are not linked by covalent bonds). Although the description of non-bonded contacts may not be that critical for accuracy of high-resolution (2 Å or better) X-ray structures, it does play a significant role in determining the limits of accuracy attainable in NMR structure determination.

There have been extensive NMR studies on the solution structure of PrP. The X-ray crystallographic approach has been limited by difficulties in crystallizing the protein(Heller et al). Like other membrane glycoproteins, PrP^{C} is extremely difficult to crystallize when glycosylated (Prusiner et al.)

Only two X-ray structures of PrP^{C} have been reported so far: the crystal structure of a truncated mutant of the Human prion protein in dimeric form (Knaus et al) and the crystal structure of residues 121-231 of the monomeric form of the C-terminal domain of the sheep prion protein (Haire et al).It has proven to be difficult to get the $PrP^{C} \rightarrow PrP^{Sc}$ conversion details as the PrP^{Sc} sample is hard to purify for biochemical and structural characterization.

The critical region we are trying to improve in this study, is the loop region in PrP^{C} (residues 167-171). This is the potential binding site for the hypothized chaperone – Protein X. This region is well-conserved in mammalian species, but is under-determined in most mammalian species due to the lack of NOE restraints. A structural quality analysis using PROCHECK, shows that none of the residues in this region fall within the most favorable regions of the Ramachandran plot. This either implies that enhancing the structure in this critical region might be crucial to elucidate the interactions of the prion protein and Protein X or alternatively, that this loop is flexible, and the NMR structure is built from data reflecting some average of these forms that might not actually be a feasible form.

NMR structures can be enhanced by adding standard peptide information such as dihedral angles as mentioned before, and/or inter-atomic distances based on statistical analysis of databases of high-resolution protein structures.

For this study, it was found better to work with inter-atomic distances as they can improve the structures with increased precision and accuracy, and can perform in an equivalent way to having same

experimental NMR restraints such as torsion angle restraints, without compromising the structure quality. Also, these constraints impose no extra cost on NMR structure refinement.

For the NMR structure refinement of the prion protein, the distributions of inter-atomic distances in known protein structures and in particular, in known X-ray structures were studied and used to extract additional distance constraints.

Inter-atomic distances can be categorized according to the specific atom pairs: residue pairs and sequential residue separation (that is, residues separated by one or more residues).

Different types of distributions are subject to different statistical distributions in the structural databases, which have been employed to construct various statistical potentials for contact determination, inverse folding, structure alignment, x-ray structure refinement etc.

In the past, Subramanium et al employed improved quantitative methods to study macromolecular ensembles for which data are scarce. They derived a database potential from distributions of inter-atomic distances obtained from a database of known structures. They reined X-ray crystal structures by molecular dynamics with the new energy function replacing the Van der Waals potential.

In our study, a large set (3900 structures) of high resolution(2 Å or more) protein X-ray crystal structures with sequence similarity of 90% or less were downloaded from the Protein Data Bank (PDB) and used to obtain the statistical distributions of distances of the afore mentioned different types.

The distances for selected pairs of atoms across one or two residues along the protein backbones (called cross-residue inter-atomic distances) are sampled to obtain the probability distribution of the distances. In order to reduce the errors in the distances and hence improve the NMR structures, the distribution functions for selected cross-residue inter-atomic distances are used to extract probable ranges for the distances.

The original constraints and derived constraints are used for structural refinement. This increases the acceptance rate in an ensemble = # of accepted structures/ # of trial structures used.

3.2 E200K variant of Human Prion Protein

3-D structures of monomeric and dimeric human prion protein (HuPrP) have been revealed by NMR spectroscopy and X-ray crystallography. However, the NMR-determined structures are not as good in quality as the X-ray structures. This limits the usefulness of the NMR structures. This issue was dealt with using a novel NMR structure refinement approach(discussed in later chapters), which is based on addition of distance constraints derived from a database of high-quality X-ray structures. Because of the importance of understanding the pathogenesis of the transformation of PrP^{C} to PrP^{Sc} that causes diseases like CJD, in this study, we chose the familial CJD-related E200K variant of the Human Prion Protein as a test case to demonstrate the usefulness of this refinement approach.

The E200K variant of the Human Prion protein contains a point mutation at the 200^{th} residue. The 200 codon is glutamate and E200K occurs just before the third helix.

Molecular dynamics simulations have been performed to study the effect of the point mutation on the dynamics of the Human prion protein. Remarkably, apart from minor differences in flexible regions, the backbone tertiary structure of the E200K variant is nearly identical to that of the wild type protein. The only major consequence of the mutation is the perturbation of the surface electrostatic potential. The present structural data strongly suggest that protein surface defects leading to abnormalities in the interaction of prion protein with auxiliary proteins/chaperones or cellular membranes should be considered key determinant of a spontaneous transition from PrP^{C} to PrP^{Sc} in the E200K form of hereditary prion disease- Creutzfeldt Jakob Disease (CJD).

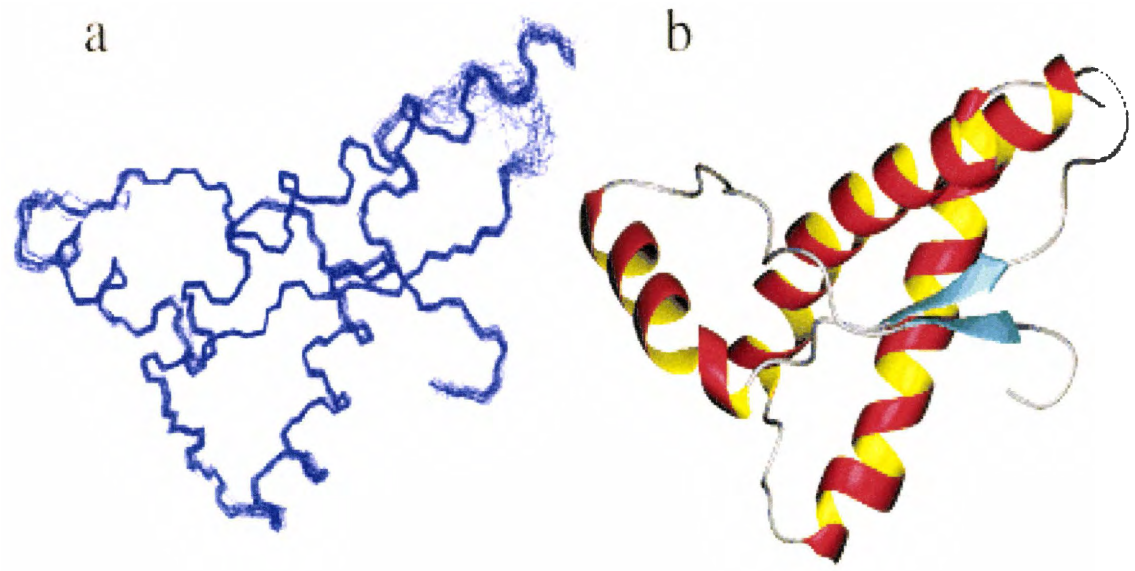


Figure 3.1: NMR Solution structure of E200K Variant of Human prion protein.

1a) 1FO7: NMR solution structure

1b) Average minimized NMR structure

Source: http://www.jbc.org/content/vol275/issue43/images/large/bc4108785002.jpeg

Comparison between E200K variant and wild-type human prion protein

Apart from minor differences in flexible regions, the backbone tertiary structure of the E200K variant is nearly identical to that reported for the wild-type human prion protein. The only major consequence of the mutation is the perturbation of surface electrostatic potential (Zhang et al). The present structural data strongly suggest that the protein surface defects leading to abnormalities in the prion protein interaction with auxiliary proteins/chaperones or cellular membranes should be considered key determinants in the spontaneous conversion of $PrP^{C} \rightarrow PrP^{Sc}$ in the E200K form of hereditary prion disease.

Note: The importance of this comparison, is because of the fact that there is no X-ray crystallography structure available for study for the E200K variant. And in our study, we had to compare it with the X-ray structure of the wild-type human prion protein.

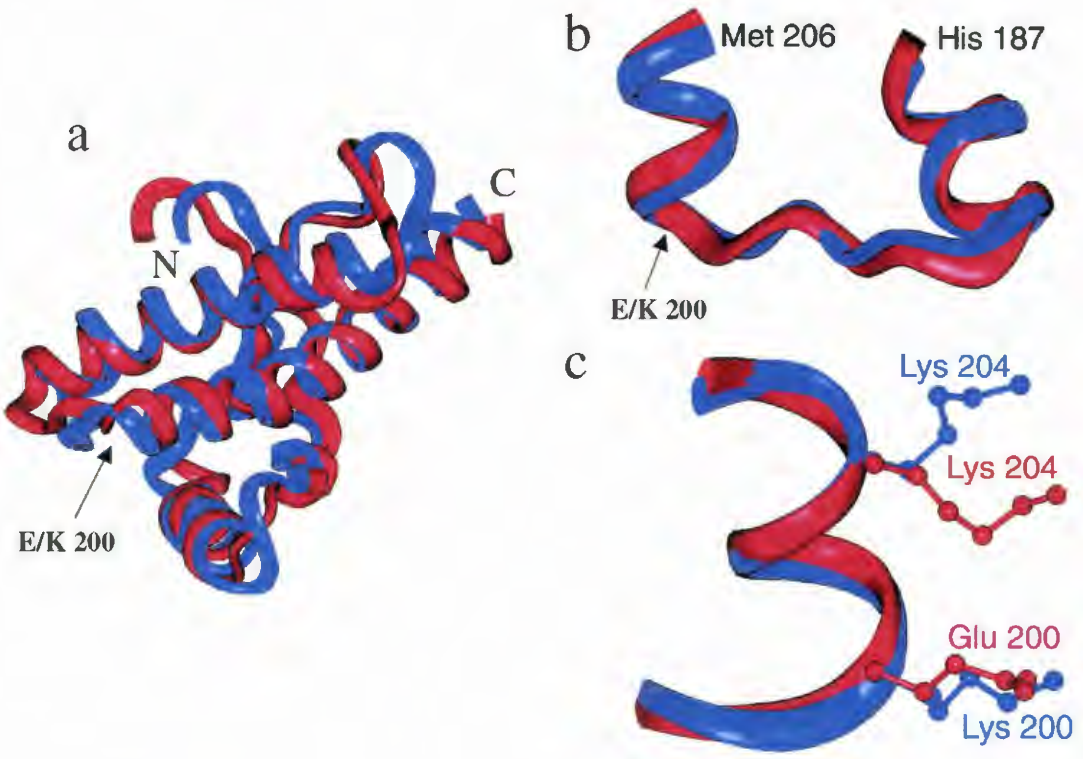


Figure 3.2. Comparison of wild type Human Prion protein(red) with E200 Variant of Human Prion protein(blue).

2(a) Ribbon presentation of the C-terminal domain based on a best-fit superposition of residues 125-228.

2(b) The best-fit superposition of the residues 187-206 shows the identical backbone conformations in this region.

2(c) Close-up view of the mutation site in both proteins, highlighting the side-chain conformation of residues 200 and 204.

Source: Journal of Biological Chemistry Vol. 275, No.43, pp 33650-33654, 2000.

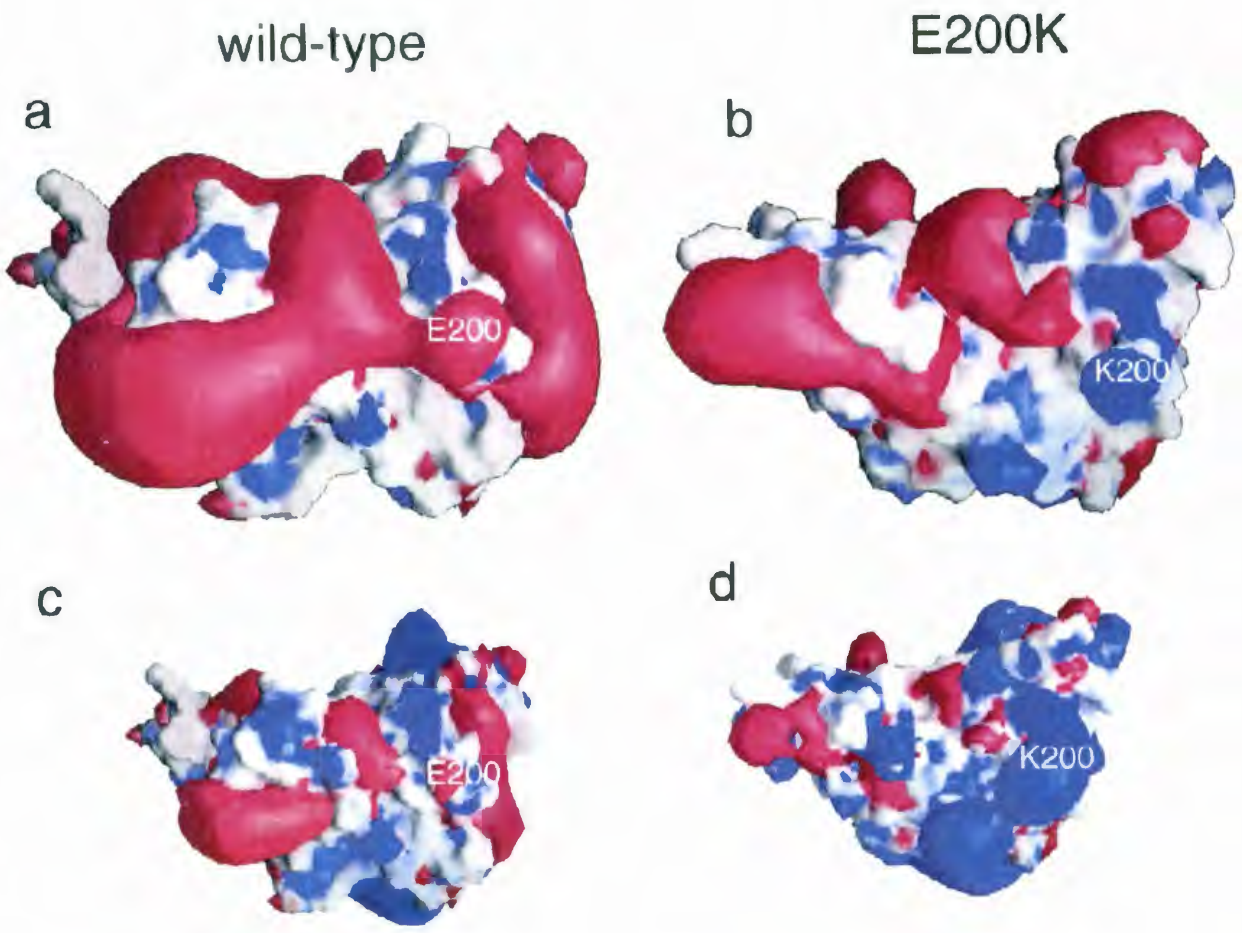


Figure 3.3 Electrostatic potential of E200K variant (right) when compared with the wild-type human prion protein(left).Blue and red colors represent positive and negative potential, respectively.

Source: Journal of Biological Chemistry Vol. 275, No.43, pp 33650-33654, 2000.

3.3 Obtaining the distribution

3900 X-ray crystal structures with resolution 2 Å or more, and sequence similarity of 90% or less, were obtained from the PDB data bank.

The residue types studied : all 20 amino acids

Atoms selected : (i) Amide N

(ii) C_α

(iii) Carbonyl C

(iv) O

(v) C_β in side-chain

The distances are specified together with the types of atoms pairs, the types of residue pairs and the sequential separations.

Consider two residues R1 and R2, separated by S number of residues. In this study, to generate minimal amount of required data for the derived constraints, S is either 0 or 1, although it could be extended to more general types of distances. Let D be the distance between Atom A1 on R1 and A2 on R2, which we want to derive.

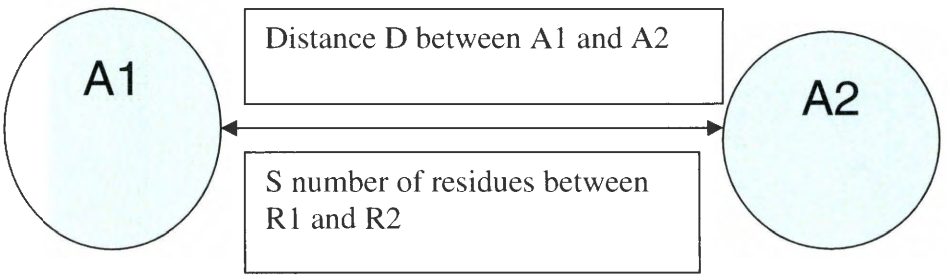


Figure 3.4: Illustration for obtaining the distance distributions.

For each set of A1, A2, R1, R2, and S, all corresponding distances in the downloaded crystal structures are calculated. The distribution of the distance D can be represented by using a probability function, P(A1,A2,R1,R2,S)(D). There are totally 20 amino acids, and 5 kinds of atoms. S can take on two possible values, 0 or 1.Therefore, number of possible distributions = 5*5*20*20*2 = 20,000.

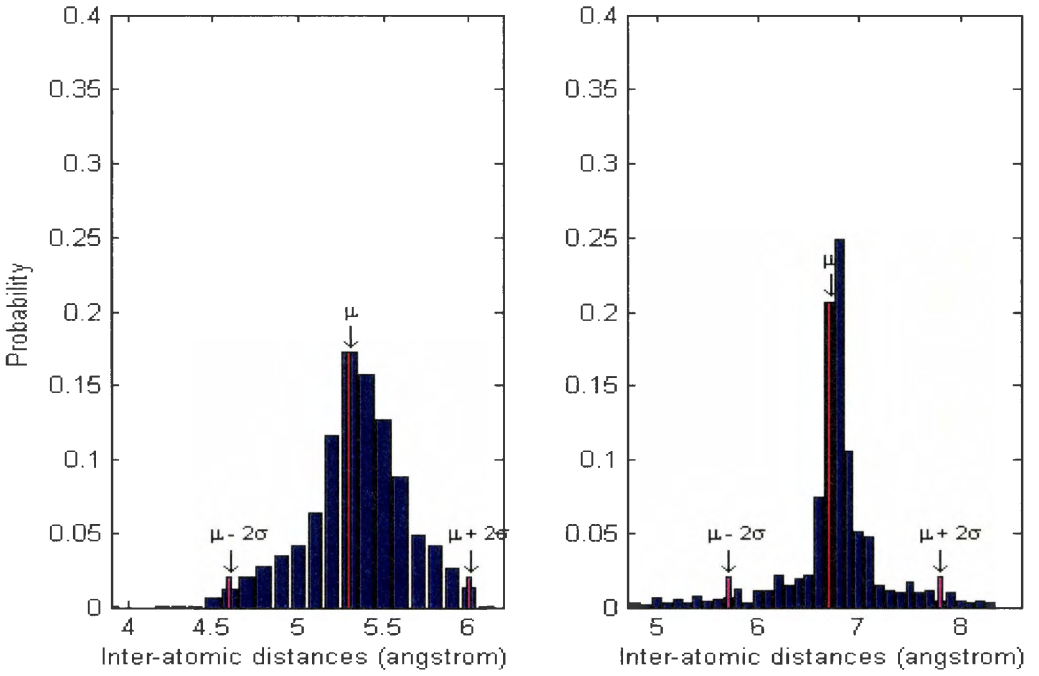
The downloaded distances are collected into bins of uniform distance intervals [Di,Di+1] where

Di =0.1* i Å i= 0,1,…,n.

Then the distribution function for any D in [Di,Di+1] is defined to be the number of distances in [Di,Di+1] normalized by the total occurrences of distances in all intervals. That is, the distribution of each type of distances was calculated by counting the number of occurrences of the distances within a set of the distance intervals.

Graph 1 illustrates two distance distributions. 1(a) shows the distances between the Cβ atom of GLU at position i and the Cβ atom of ASP at adjacent position i+1.

1(b) is for distances between Cβ atom of ALA at position i and the Cα atom of GLU at the next nearest neighbor at position i+2.



Graph 3.1: Probability distribution of Inter-atomic distances of high-resolution structures

These graphs show clearly the non-uniform distribution of distances. 200 distance intervals, of length 0.1 A each, are specified in the horizontal axis. The ordinate values show the frequencies of the distances in the corresponding distance intervals. The means μ and standard deviations σ of the distributions are used to specify the range constraints to be $\mu - 2\sigma$ and $\mu + 2\sigma$ for the corresponding distances.

The precision of an ensemble is usually calculated in terms of Root Mean Square Deviation(RMSD). But this can be over-estimated as current modeling software may not necessarily contain the whole range of structures determined b the distance constraints. There are loops and tails in a folded protein structure where NMR experimental data cannot define the structure well.

The problems still associated with the above theoretical method, is that while some of the deviations can be attributed to additional flexibility of the protein in solution, many of them must originate in modeling errors. Unfortunately, there is no clear way of distinguishing between the two.

Large deviations of inter-atomic distances in NMR structures from their average distributions in known protein structures are clear indications of modeling errors, possibly due to the lack of proper constrains on the corresponding distances in the NMR data. This issue was dealt with by using the novel approach of confining the distances to their most probable ranges according to their distributions in known protein structures. The distance distribution functions are used to generate a set of lower and upper bounds that act as the new (derived) constraints.

Lower bound = mean + 2 * (standard deviation)

Upper bound = mean - 2 * (standard deviation)

The generated distance bounds are then taken as additional distance constraints to refine a set of NMR structures (in our case, the NMR solution structure of the Human Prion variant E200K, PDBID: 1FO7 ; the X-ray structure for the Human prion protein is available, PDBID: 1I4M).

Two critical under-determined loop regions were targeted for improvement in this study: residues 167-171 and loop residues 195-199.

NMR experimental data (original constraints)for the h PrP^{C} C-terminal globular domain (residues 125-228) (see Table 1) was downloaded from BioMagResBank (BMRB).

Table 3.1 **Experimental restraints**

Experimental Restraints in Loop1 (residue 167-171)

Residue	NOE[†]	Torsion	H-bond	J- coupling[‡]
166	27	0	0	0
167	3	0	0	0
168	3	0	0	0
169	0	0	0	0
170	1	0	0	0
171	21	1	2	0
172	17	1	0	0

Experimental Restraints in Loop2 (residue 195-199)

Residue	NOE	Torsion	H-bond	J-coupling
194	35	1	0	0
195	17	0	0	0
196	45	0	0	1
197	34	0	0	0
198	76	0	2	1
199	32	1	0	1
200	27	1	2	1

	NOE	Torsion	H-bond	J-coupling
Total	3157	177	96	44
Per Res.	29.8	1.7	0.9	0.4

Note: Total number of Residues = 106.

†Total distance restraints, ‡J HNHA-coupling constants

The structure of this domain was then refined using NMR experimental constraints and the additional database-derived constraints, by implementing the standard torsion angle dynamic simulated annealing protocol implemented in CNS.

The results obtained with and without additional database-derived distance constraints are examined on the deviations of selected inter-atomic distances from their average distributions, and are compared and assessed in terms of several criteria used in NMR modeling:

i. acceptance rates of the structures in the ensemble: The acceptance rate for the ensemble of structures is defined as the number of accepted structures divided by the total number of trial structures, including the "rejected" ones. A trial structure is accepted if it meets all the default acceptance criteria in CNS (including bond lengths, bond angles, NOW distances, dihedral angle restraints)

ii. Root Mean Square Deviation (RMSD) values of the ensemble of structures : The RMSD is a measure of the precision of an ensemble.

iii. RMSD values of the structures compared with their X-ray structures.

iv. Ramachandran plot analysis

v. PROCHECK summary statistics and analysis.

It was observed (Feng Cui et al) that with additional database derived constraints, the acceptance rates of the refined NMR structures increases, indicating that the addition of these constraints may not only aid in

correcting the distance errors in the NMR structures, but may also improve the performance of the modeling program for obtaining acceptable ensembles of structures.

While a distance constraint can be derived for every selected pair of cross-residue atoms based on the distribution of the distance in known protein structures, not all the constraints are necessary for the refinement of a given NMR structure since some distances may not necessarily be incorrect even if they deviate significantly from their average distributions. In other words, we can use a sparse set of constraints, thereby reducing the computational and analysis load.

The constraints may be most effective for distances or interactions in regions that are under-determined by NMR experimental data (for instance- the loop regions of 1FO7).

On the other hand, the atom types can certainly be extended to include more side-chain atoms and longer-range interactions. In general, the backbone and other non-hydrogen atoms are perhaps most likely to have distances among them disagreeing with their distributions, since the non-hydrogen atoms usually do not have as much distance data available as hydrogen atoms and therefore cannot be determined as directly and accurately.

3.4 Refinement of E200K variant of the Human Prion Protein

Three ensembles of 1FO7 NMR structures were generated (50 structures in each ensemble) for the residues 125-231; selection done on the basis of lowest energy criteria:

Ensemble 1: 50 accepted structures of 1FO7 generated only from NMR experimental data.

Ensemble 2: 50 accepted structures of 1FO7 generated from both experimental data and a set of database-derived distance constraints which includes:

constraints C-C_β and N-C_α , where C and N are in a residue on position *i* and C_β and C_α are in a residue on position *i*+1

Ensemble 3: 50 accepted structures of 1FO7 generated from both experimental data and a set of database-derived constraints which includes:

a. constraints C_β - C_β, C-C, N- C_α and C_α –C, where C, N, C_α and C_β are in a residue on position *i* and C, C_α, C and C_β are in a residue on position *i*+1.

b. constraints C_β - C_β, where the first C_β is in a residue on position *i* and the second one is in a residue on position *i*+2.

c. constraints C_β - C_β , where the first C_β is in a residue on position *i* and the second one is in a residue on position *i*+2.

Based on the distribution, we can easily get the mean and standard deviation of the distribution. We use (mean - 2 * std) as the lower bound of the inter-atomic distance for any two atoms, while (mean + 2 * std) as the upper bound of this distance. Therefore we can decide the range for this distance and use this range as an additional constraint for refining NMR structure.

NMR experimental data (These were downloaded from BMRB site) for 1FO7 includes –

i. restraints for torsion angle

ii. restraints for hydrogen bond

iii. distance restraints derived from the Nuclear Over Hauser Effect (NOE)

iv. Additional distance constraints that are based on a database of X-ray structures were also generated.

CNS 0.9 was used for structure calculation with standard torsion dynamics simulated annealing protocol.

The program suite PROCHECK was used for:

-analyzing the geometry

-analyzing the restraint violations of the average and minimized structure of each ensemble.

PROCHECK outputs include:

-Various plots in PostScript format

-Summary statistics.

The minimal energy structures of each of the ensembles are evaluated in terms of their agreement with the experimental constraints and optimal covalent geometry, their local and global potential energy, and the overall structure quality.

Two of the ensembles were analyzed. The first ensemble, without added constraints-$\langle SA \rangle^{NMR}$ and Ensemble 3, with added database derived constraints -$\langle SA \rangle^{NMR+D}$. $\langle SA \rangle^{aNMR}$ and $\langle SA \rangle^{aNMR+D}$ represent the average and energy minimized structures of the respective ensembles. Ensemble 2 was not included in the analyses as we observed better improvement in Ensemble 3.

Table 2 summarizes that the average root mean square deviation (RMSD) of the structures in the ensembles as well as RMSDs of minimal energy structures in the ensembles from the experimentally specified constraints and optimal covalent geometry are comparable for both $\langle SA \rangle^{NMR}$ and $\langle SA \rangle^{NMR+D}$.

Table 3.2: Ensemble Analysis of Structures Refined With and Without Database-Derived Constraints

Mean r.m.s. deviations from experimental	$\langle SA \rangle^{NMR}$	$\langle SA \rangle^{aNMR}$	$\langle SA \rangle^{NMR+D}$	$\langle SA \rangle^{aNMR+D}$
distance restraints (Å)	0.0046 ± 0.0018	0.0030	0.0047 ± 0.0016	0.0040
dihedral restraints (degrees)	0.1664 ± 0.0368	0.1540	0.1589 ± 0.0340	0.1380
J-coupling restraints (Hz)	0.3787 ± 0.0951	0.2470	0.2105 ± 0.0186	0.2550
Mean r.m.s. deviations from idealized				
Bond geometry (Å)	0.0014 ± 0.00021	0.0012	0.0014 ± 0.00024	0.0012
angle geometry	0.3128 ±	0.2990	0.3108 ±	0.3040

(degrees)	0.0212		0.0153	
improper geometry (degrees)	0.2148 ± 0.0236	0.2000	0.2105 ± 0.0186	0.2120
Energy (*kcal/mol*)				
Total	104.16 ± 24.80	82.08	102.31 ± 23.09	86.30
Bonds	3.30 ± 1.11	2.45	3.54 ± 1.58	2.70
Angles	46.78 ± 6.67	42.53	46.11 ± 4.92	44.11
Improper	6.78 ± 1.54	5.80	6.49 ± 1.21	6.53
Van der Waals	34.44 ± 9.44	26.23	31.81 ± 6.97	26.29
NOE	5.85 ± 4.80	2.11	6.97 ± 6.57	3.60
Dihedral	0.31 ± 0.14	0.26	0.28 ± 0.14	0.21
Measures of structure quality				
Residues in most favorable region	85.40%	84.40%	89.60%	88.50%

Residues in additional allowed region	14.60%	14.60%	10.40%	11.50%
Residues in generously allowed region	0.00%	0.00%	0.00%	0.00%
Residues in disallowed regions	0.00%	1.00%	0.00%	0.00%

For $<SA>^{NMR+D}$, the total energy and energy of improper angle restraints and dihedral angle restraints are lower than that of $<SA>^{NMR}$, although not by so much.

PROCHECK results on average and energy-minimized structures and minimal energy structures of both ensembles show a significantly higher percentage (89.6%) of residues in most favorable regions of the Ramachandran plots of the structures in $<SA>^{NMR+D}$.

For $<SA>^{NMR}$, the percentage of these structures in the most favorable regions is 85.4%. Note here that, this percentage for $<SA>^{NMR}$ is consistent with what was reported by Zhang et al. for their experimental results (85.7%).

The increase in the percentage of residues in the most favorable regions shows a clear improvement of the structures. As the increase is seen in both the minimal energy structure and the average energy-minimized structure, it is evident that the improvement has occurred overall through out the ensemble of structures. However, previously, such a high percentage of residues in most favorable regions of the Ramachandran plot were observed only in the minimal energy structure of a structural ensemble but not in both the average and energy-minimized structure and the minimal energy structure.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Comparative Analysis

Back-bone structure superimposition and side-chain packing of loop regions

Differences between the average and energy minimized structures $<SA>^{aNMR}$ and $<SA>^{aNMR+D}$ in their (φ,ψ) angle values in the loops reflect the different loop conformations between the two structures (As indicated in Figures 1, 2). As shown in Figure 1, $<SA>^{aNMR+D}$ (magenta square) and $<SA>^{aNMR+D}$ (green square) are extremely close to each other, except at the two loop regions (residues 167-171 and 195-199).

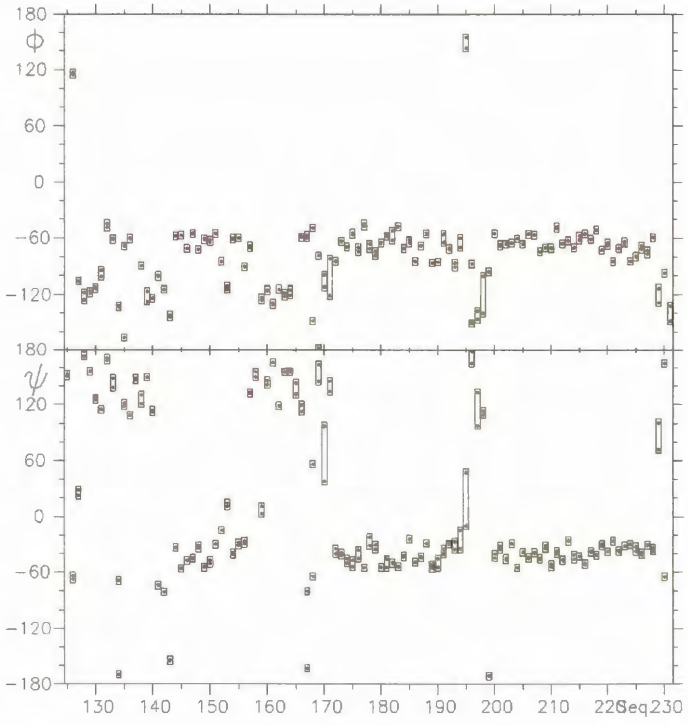


Figure 4.1: Plots of Psi and Phi Angles in Representative Structures Refined With and Without Database-Derived Distances

Ramachandran Plot Analysis

To further elucidate the improvement in the under-determined loop regions of the E200K variant of the Human prion protein, the sequential φ and ψ angles for each residue of $<SA>^{NMR}$ and $<SA>^{NMR+D}$ are compared (refer to Figure 1) and also displayed in Ramachandran plots (see Figures 2) and analyzed.

In the loop region between helix 2 and helix 3 (residues 195-199) the angles φ and ψ of $<SA>^{aNMR+D}$ are closer to the mean structure of the 30 best structures of 1F07 reported previously by Zhang et al., than the φ and ψ angles of $<SA>^{aNMR}$. The 30 best structures were selected from 60 calculated structures, which were believed to be most accurate. In contrast, the residues of $<SA>^{NMR}$ in the loop region (residues 167-171) lie far outside the favorable regions of the Ramachandran plot (Figure 2a), and so do the residues of the same loop in the mean structure of the 30 best structures (1F07)and average and energy-minimized structure (1FKC) of the ensemble generated by Zhang et al. However, after refinement incorporating the database-derived distanced constraints, that is, in the average and energy-minimized structure of $<SA>^{NMR+D}$, most of the residues move into the most favorable regions of the Ramachandran plot (Figure 2b). These are small yet significant changes.

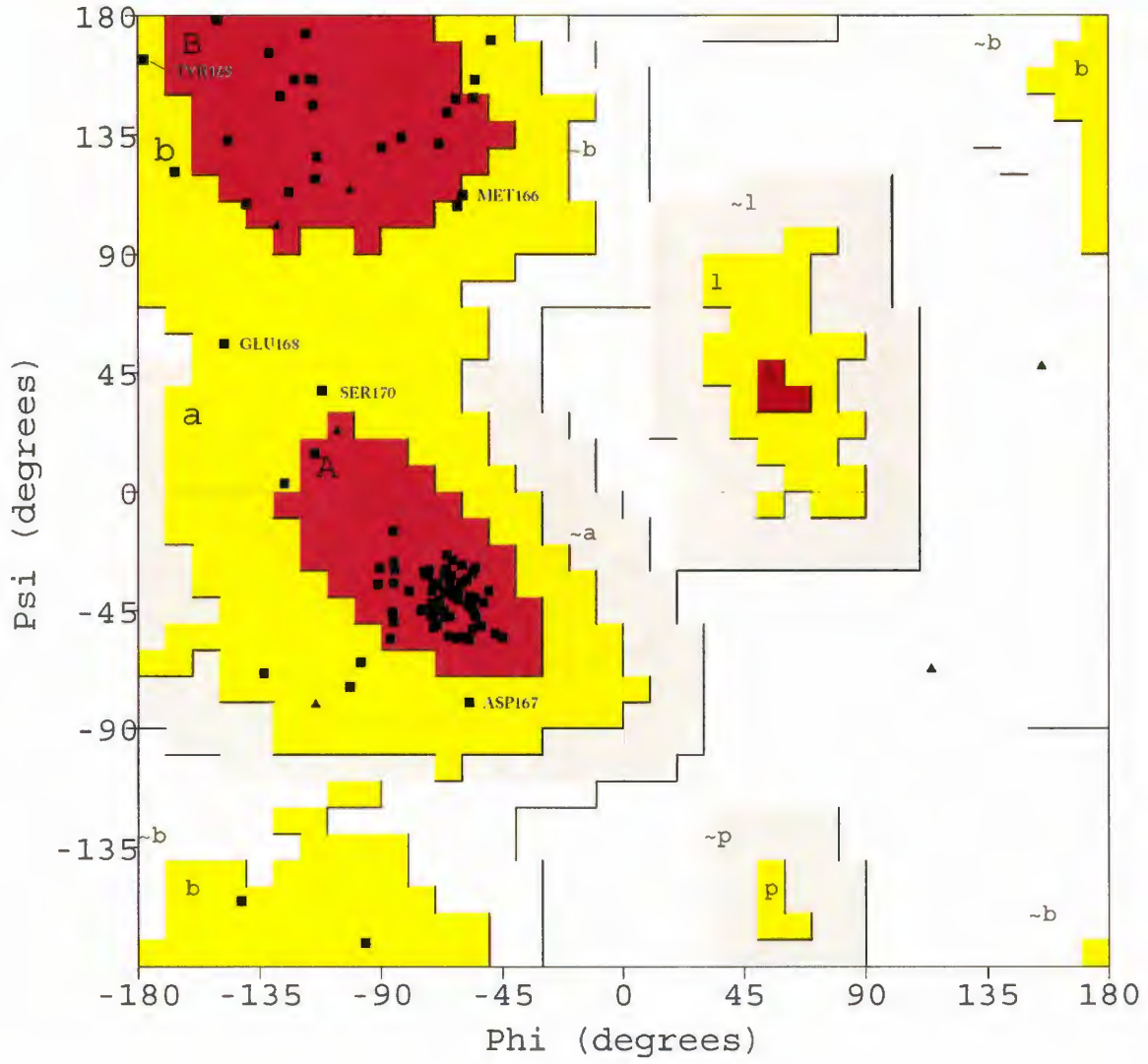


Figure 4.2(a): Ramachandran Plot of Representative Structure Refined Without Database-Derived Constraints

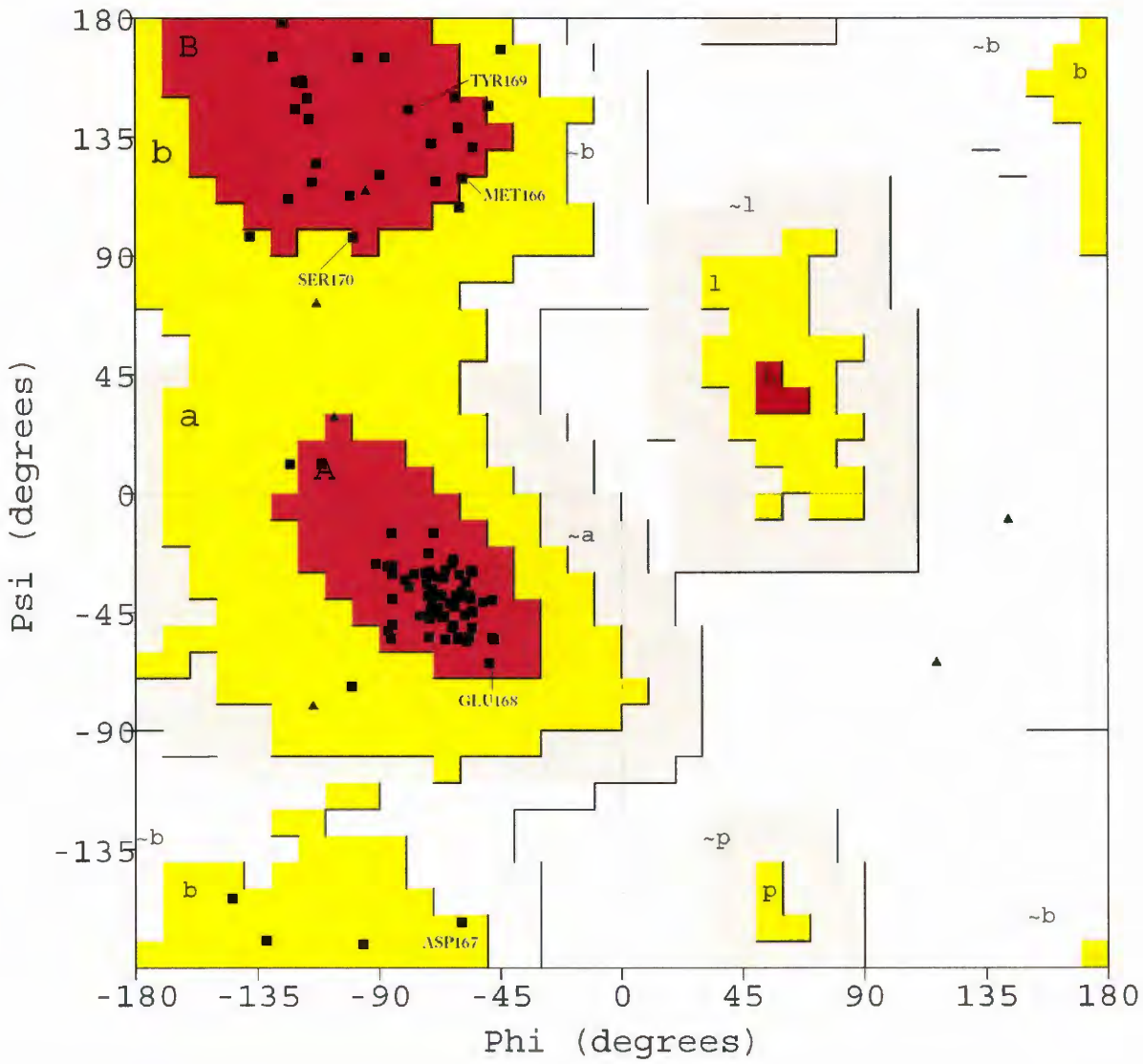


Figure 4.2(b): Ramachandran Plot of Representative Structure Refined With Database-Derived Constraints

Back-bone structure superimposition

The backbones of the two structures $<SA>^{NMR}$ and $<SA>^{NMR+D}$ can be superimposed in 3-D, as shown in green and magenta cylinders for $<SA>^{NMR}$ and $<SA>^{NMR+D}$ respectively in Figure 3.Figures 4 and 5 show the backbones of the loop regions in more detail.

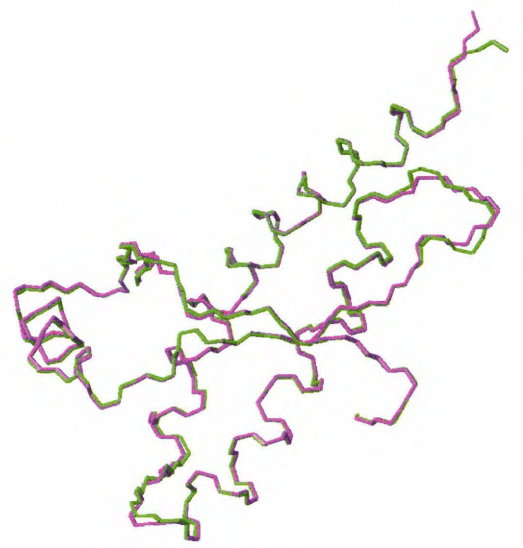


Figure 4.3: Superimposition of Representative Structures Refined With and Without Database-Derived Constraints;Legends: Green - $<SA>^{aNMR}$, Magenta - $<SA>^{aNMR+D}$

The loop (residues 167-171) in $<SA>^{NMR+D}$ appears quite different from the corresponding region in $<SA>^{NMR}$ (Figure 4). This implies that the database-derived distance constraints can actually affect the backbone conformations in regions where experimental restraints are insufficient (Table 1), and in other words, these new constraints applied here do not exert their influence uniformly through the structure, but rather in a localized way to improve the most under-determined parts.

The conformations of another loop (residues 195-199) in both structures appear quite similar except at residue GLY195 (Figure 5). It has been shown that the conformations of glycine and neighbouring residues can indeed be improved by using database-derived distance constraints.

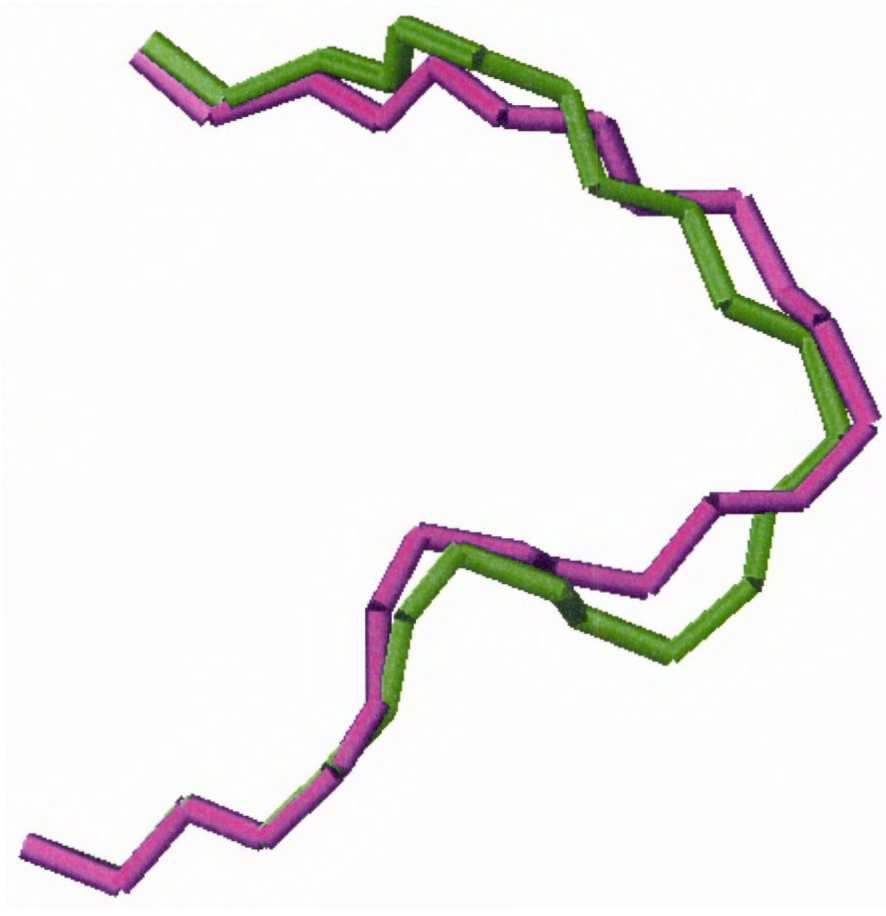


Figure 4.4: Superimpositions of backbones of Loop 1 (167-171)

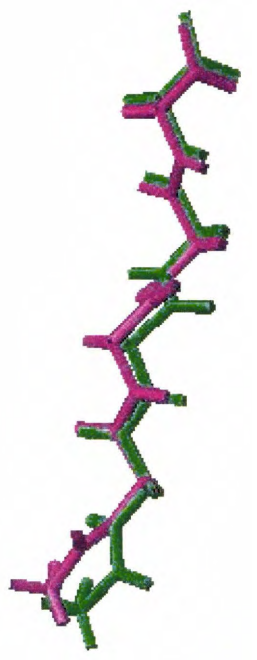


Figure 4.5: Superimposition of Backbones of Loop 2 (195-199)

Examination of side-chain packing in the two loop regions show that overall conformations of side-chains are quite similar between $<SA>^{NMR+D}$ and $<SA>^{NMR}$ (Figures 6, 7). No change in either hydrogen bonds or salt bridges was observed. It suggests that the database-derived distance constraints do not particularly affect the side-chain packing in general, which is not so surprising since only the constraints between backbone atoms have been utilized in this study. The impact of introducing further distance constraints in the refinement process could be substantial and has been left as a suggestion for future research work.

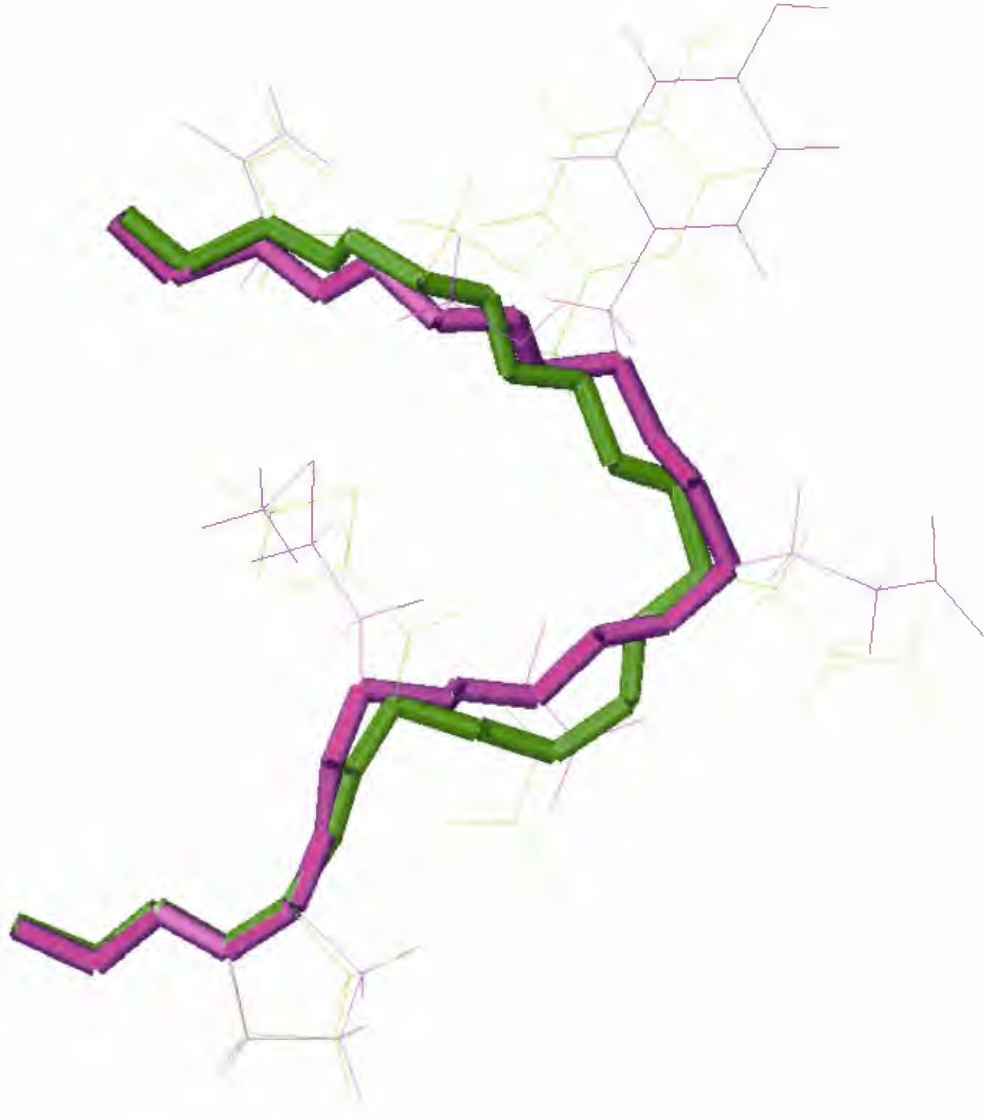


Figure 4.6: Superimposition of backbones and side-chains of Loop 1

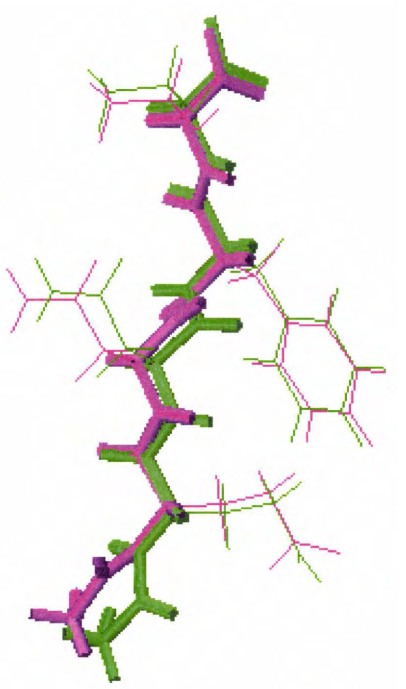


Figure 4.7: Superimposition of backbones and side-chains of Loop 2

NMR and X-ray structure comparison for PrP^C wild-type

The C-terminal domain (residues 125-231) of the prion protein, has been previously determined by NMR under neutral (PDB ID: 1HJM, ph 7.0) and under mildly acidic (PDB ID: 1QM0, pH 4.5) conditions as well as by X-ray crystallography (1I4M for a dimeric form of $HuPrP^C$ and 1UW3 for a monomeric form of $shPrP^C$).

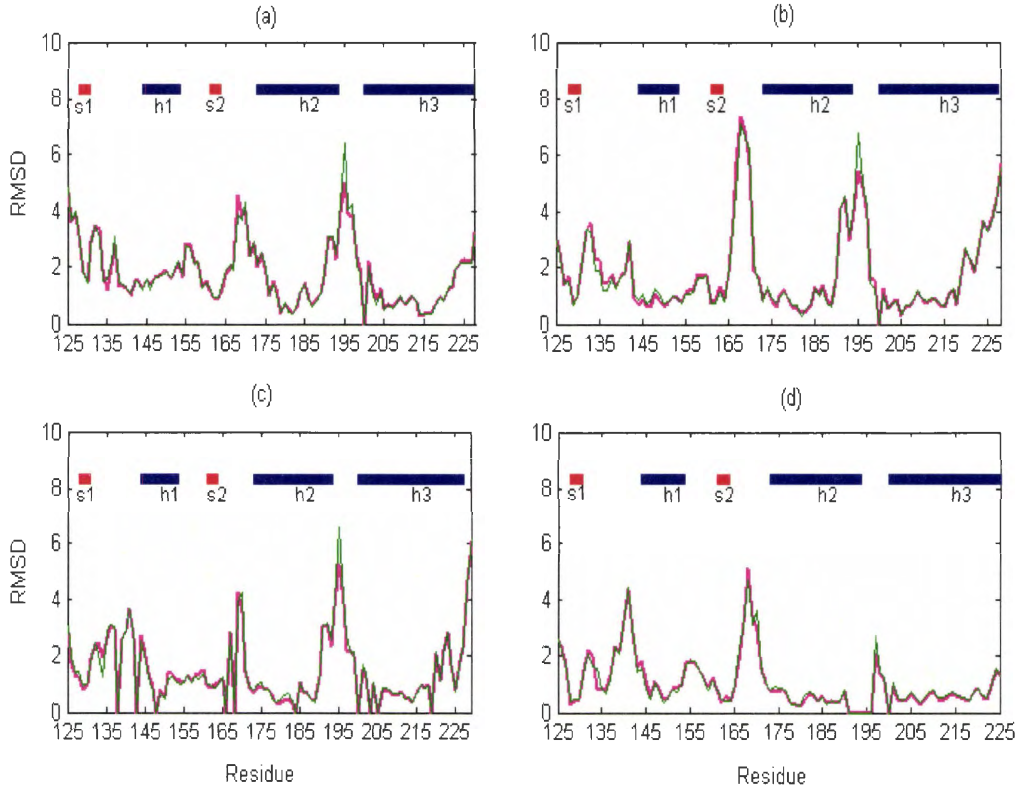
Residue-residue comparisons of the average and energy-minimized structures, $<SA>^{aNMR+D}$ and $<SA>^{aNMR}$, with the NMR and X-ray structures (also average and energy-minimized) of the PrP^C wild types were conducted to measure the improvements in the accuracy of the structures, especially the under-determined loop regions, after the database-derived distance constraints were used for refinement.

The residue RMSD values for the average and energy-minimized structures (calculated for the backbone atoms, N, C^{α}, C', and O) of $<SA>^{NMR+D}$ and $<SA>^{NMR}$ compared with 1QM0, 1HJM, 1UW3, and 1I4M, respectively, are plotted in Graphs 2, with magenta for $<SA>^{aNMR+D}$ and green for $<SA>^{aNMR}$ in each of the plots.

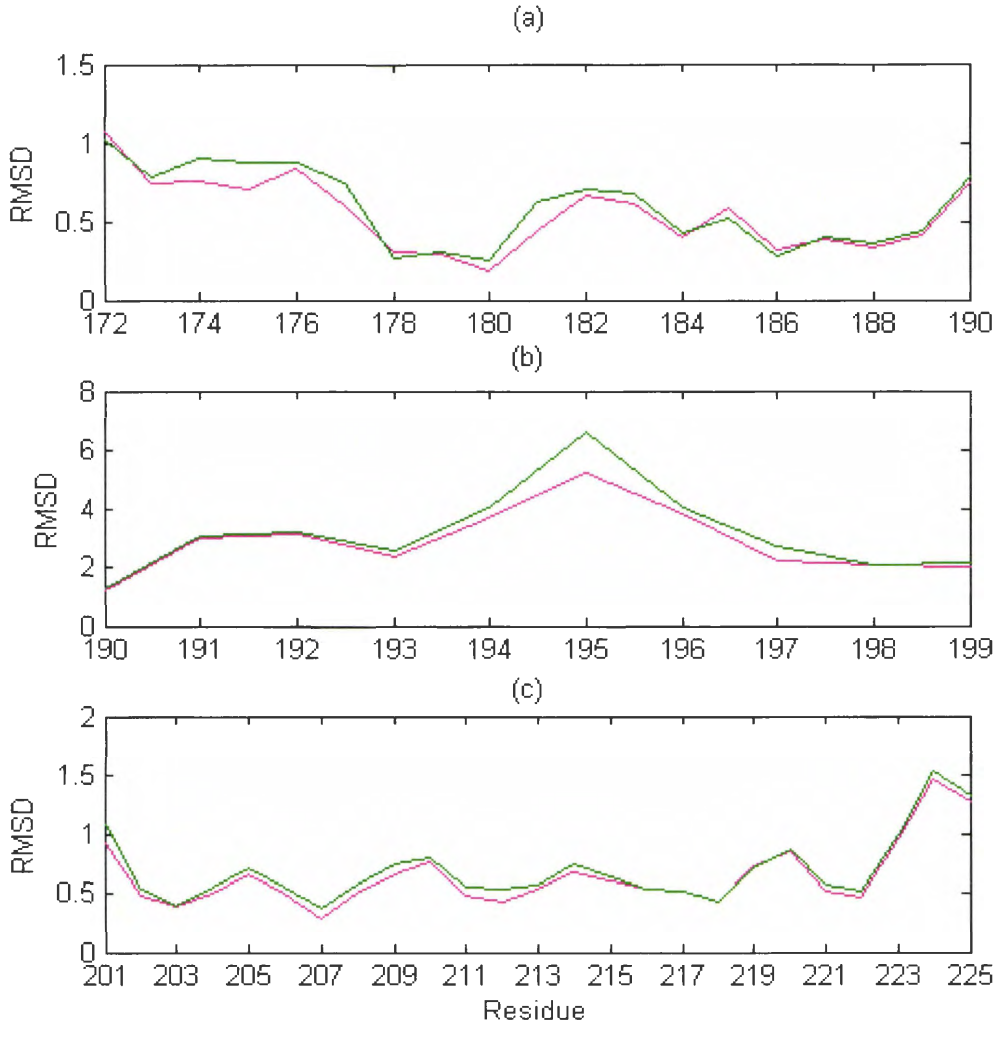
The secondary structures are indicated along the top of each part of the figure with *h* representing an alpha helix and *s* a beta sheet. The residue RMSD values in the loop regions (residues 167-171 and 195-199) in all these plots are all significantly higher than in the remainder of the structure (> 4 Å), which suggests that the loop regions are relatively more flexible. The helix regions seem more rigid (residue RMSD values around 2Å). However, in the loop region between the helix 2 and helix 3 (residues 195-199), the residue RMSD values for $<SA>^{NMR+D}$ are consistently smaller than those for $<SA>^{NMR}$ (see Figure 1a-d), showing that the database-derived distance constraints modify this loop to be more consistent with the NMR and X-ray structures of other prion variants. Indeed, the loop was poorly determined in the case of $<SA>^{NMR}$, especially around GLY195, mainly due to insufficient NMR data in the region (refer to Table 1), but was refined by the introduction of additional database-derived distance constraints.

In the other loop, between the β-sheet and α-helix 2 (residues 167-171), $<SA>^{NMR+D}$ appeared closer to the $HuPrP^C$ X-ray structure (1I4M) with smaller residue RMSD values than $<SA>^{NMR}$ (see Graph 2d), although it is not so close to other wide types.

In addition to the under-determined loop regions, differences between $<SA>^{NMR+D}$ and $<SA>^{NMR}$ were observed in well-defined helix regions (helix 2 and helix 3) as well, as shown in Graph 3, where a monomeric form of $HuPrP^{C}$ X-ray structure was used as a reference structure because it could be superimposed on the refined NMR structures, especially around the residues 125-190 and 197-225. The residue RMSD values for $<SA>^{NMR+D}$ are slightly smaller than those for $<SA>^{NMR}$ in the N terminal of helix 2 (residues 172-190) and helix 3 (residues 201-228), showing that the helix regions of $<SA>^{NMR+D}$ are nearer to the corresponding X-ray structure than those of $<SA>^{NMR}$ (refer to Graph 3a, 3c). In the loop region between the β-sheet and helix 2 (residues 191-199), $<SA>^{NMR+D}$ is closer to the sheep PrP^{C} X-ray structure than $<SA>^{NMR}$ (refer to Graph 3b) with smaller residue RMSD values. Here, the human PrP^{C} X-ray structure was not used as the reference structure for this region because it is a switch region connecting the helix 2 to the swapped helix 3 in the $HuPrP^{C}$ X-ray structure. Overall, $<SA>^{NMR+D}$ has slightly better agreement than $<SA>^{NMR}$ in both the under-determined loop regions and in well-defined helix regions, when compared against the NMR and X-ray structures of other PrP^{C} variants.



Graph 4.1: **Residue-Residue RMSD Comparison**



Graph 4.2: **Detailed RMSD Comparison in Helix 2, Loop 2 and Helix 3**

4.2 Results and discussion

Results

Many prion diseases are linked with point mutations in the Prion coding gene (PRNP). More than 20 mutations in PRNP have been associated with prion diseases. The E200K variant of the Human prion protein, which results from the point mutation where glutamic acid is substituted by lysine at the 200^{th} residue, is a major cause of familial CJD.

The tertiary structure of the E200K variant is almost identical to that of the wild-type Human prion protein (except for a few minor differences in the flexible regions). This mutation not only changes the surface electrostatic potential of the protein which in turn may affect its interaction with Protein X and other cofactors or cellular components, but also the stability of the prion protein. Yet, by itself, the mutation cannot lead to the $PrP^{C} \rightarrow PrP^{Sc}$ conversion, which could require additional modification of the protein and involvement of other cofactors. Due to the hypothetical involvement of Protein X in the pathogenesis of familial CJD, it is of considerable interest to investigate the structure of the binding sites between the E200K variant and Protein X, and refine any such areas which are under-determined. This includes the loop regions encompassing residues 167-171 and residues 215 and 218.

In order to refine the conformation of this critical region, we employed distance constraints that are derived from a large database of high-resolution protein structures.

The results and analysis show that the loop region as well as the overall structure of the E200K variant of the Human Prion protein was both significantly improved by several comparisons.

On refinement with the additional database derived constraints, both the loop regions in $\langle SA \rangle^{NMR+D}$ showed more generally acceptable conformations. Without experimental data, it is rather difficult to say whether the calculated conformation reflects the 'true' conformation of the protein in solution or not. However, on comparing various NMR and X-ray structures of the wild-type prion protein, a convergence between the structures was confirmed. This implied an increase in the accuracy of the structure and in particular of the loop regions where the experimental data was insufficient.

The refined structure of the protein should provide a better understand of the structure and properties of the prion protein and thus could facilitate our insights in the conversion process in prion diseases. This novel idea could also be used to approach other diseases which are accompanied by neural accumulation of misfolded proteins, such as Alzheimer's and Parkinson's disease.

Conclusion

We extended refinement methods to improve the structures of proteins with under-determined regions, to implementing database derived constraints during the process of refinement.

Other ideas have been implemented in the past. For instance, the human thioredoxin was refined using a database-derived dihedral potential. The method showed improvement in the structure quality on the basis of structural assessment using XPLOR.

In comparison with this method, the method of applying database-derived distance constraints seemed to increase the accuracy more efficiently without compromising the quality of the model of the protein. Nonetheless, it may be of interest to consider including some form of such a dihedral constraint to evaluate the loop regions further. It may be that this will not prove significantly useful if the loop regions are highly flexible. If the region is more flexible, one can expect more variation in dihedral angles. In a corrected determined structure, you wouldn't be surprised to find more angles falling outside (perhaps only slightly in some cases) the most favorable regions. For a given residue, there is an energy strain associated with dihedral angles falling outside the 'most favorable' regions of the Ramachandran plots. But one must remember that the most favorable regions were originally defined based on the analysis of overall structural data. So a conformation may be globally favorable if the dihedral angle strain is offset by tertiary interactions such as hydrogen bonding or hydrophobic interactions. The work done, could perhaps be improved further, if we give more consideration to the secondary structure of any biological molecule.

Another thing to be mentioned here, is that we had generated two Ensembles using database-derived distance constraints- Ensemble 2 and Ensemble 3 (refer to methodology). We observed improvement in the model in both ensembles, but the most significant improvement in the loop regions was observed in the average

and energy minimized structure of the third ensemble $<SA>^{aNMR+D}$. We did not have any correlation or justification for the choice of the distance constraints data set that would yield significant results. Perhaps, if one looked into further stereochemical properties of residues and atoms, it would prove helpful to investigate if there could be any correlation factor between the data set chosen and the improvement seen. Thus, then one would be able to avoid the tedious trial-and-error routine to find the best model for the protein structure under consideration.

Further, it may even be of considerable interest to see the improvement in the structure if we incorporated the dihedral potential and the distance constraints together, perhaps by relaxing some of the constraints, in order to find a energy minimum.

My hope is that, this method will enable us to get a better insight in the understanding of the prion protein structure, and thus reveal the mysteries involved in the transition to the scrapie form in prion diseases.

APPENDIX

Crystallography and NMR System (CNS)

A new software suite called Crystallography and NMR System (CNS) has been developed for macromolecular structure determination by X-ray crystallography or solution nuclear magnetic resonance (NMR) spectroscopy.

The architecture of this suite is highly flexible, allowing for extension to other structure-determination methods (like solid-state NMR spectroscopy).

CNS has a hierarchical structure: a high-level hypertext markup language (HTML) user interface, task-oriented user input files, module files, a symbolic structure determination language (CNS language), and a low-level source code which is written in FORTAN77 for UNIX-based operating systems. Each layer is accessible to the user, and source-code modification is possible. The CNS language allows the user to perform operations on data structures, such as structure factors, atomic properties etc.

User-friendly task-oriented input files are available for nearly all aspects of macromolecular structure determination by X-ray crystallography and solution NMR.

Goals of CNS

-To create a flexible computational framework for exploration of new approaches to structure determination

-To provide tools for structure solution of difficult or large structures

-To develop models for analyzing structural and dynamical properties of macromolecules

-To integrate all sources of information into all stages of the structure determination process.

To meet these goals, algorithms were moved from the source code into a high-level structure-determination language, which allows symbolic target functions, data structures, procedures and modules. The FORTAN77 code of the CNS package acts as an interpreter for the language and includes hard-wired functions for efficient processing of computing-intensive tasks. A few extensions to standard FORTRAN are used, which are converted to standard FORTAN77 code using a pre-processor. The source code consists of a highly modular

set of subroutines and functions. The main data structures reside in separate common blocks. Dynamic memory allocation is accomplished by use of the C-function 'malloc'. Installation and compilation of the program has been automated by the use of the UNIX 'make' facility. CNS has version control, i.e. the consistency of the version numbers of the task and module files is checked against the version of the executing CNS program. The CNS language provides a common framework for nearly all computational procedures of structure determination, with symbolic data structure manipulation. A comprehensive set of crystallographic procedures for phasing, density modification and refinement has been implemented in this language. User-friendly input files which can be accessed through an HTML graphical interface, are available to carry out these procedures.

CNS Capabilities

Experimental Phasing

- -heavy atom searches
- -Patterson refinement
- -multiple-isomorphous replacement phasing and site refinement
- -multi-wavelength anomalous dispersion phasing and sire refinement

Molecular Replacement

- -Patterson real-space and direct rotation searches
- -Patterson-correlation refinement
- -fast FFT- translation search

Density Modification

- -creation of envelopes
- -solvent-flattening
- -density averaging
- -histogram matching

Refinement

- -maximum likelihood targets

-torsion-angle molecular dynamics

-Cartesian molecular dynamics

-conjugate gradient minimization

-composite annealed omit map

NMR Structure Calculation

-Nuclear Over Hauser Effect (NOE)-derived distance restraints

-NOE-intensity restraints

-1-bond and 3-bond J-coupling data

-α, β carbon and proton chemical shifts

-residual dipolar coupling restraints

-diffusion anisotropy restraints

-dihedral angle restraints

-hydrogen-bond distance restraints

-simulated annealing structure calculation

-refinement

Other

-correlated dihedral angle probability conformational database

-Protein Data Bank (PDB) deposition file generation

-mmCIF file creation

There are no pre-defined reciprocal or real-space arrays in CNS. Dynamic memory allocation allows one to carry out operations on arbitrarily large data sets with many individual entries without the need for re-compilation of the source-code. The various reciprocal structure-factor arrays must, therefore, be declared and their type specified prior to invocation.

CNS data elements

CNS supports two types of data elements which may be used to store and retrieve information. *Symbols* are typed variables, such as numbers, character strings of restricted length, and logical variables. They are denoted by the dollar ($) sign. *Parameters* are untyped data elements of arbitrary length that may contain collection of CNS commands, numbers, strings, or symbols. They are denoted by the ampersand (&) sign. Symbols and Parameters may contain a single data element, or they may represent a compound data structure of arbitrary complexity. The hierarchy of these data structures is denoted using a period (.). The information stored in the data elements can be retrieved by simply referring to them within a CNS command: the symbol or parameter name is substituted by its content. Symbol substitution of portions of the compound names allows one to carry out conditional and iterative operations on such data structures, such as matrix multiplication.

CNS Modules and Procedures

Modules exist as separate files and contain collections of CNS commands related to a particular task. Whereas, *procedures* can be defined and invoked from within any file. Modules and procedures make it possible to write programs in CNS language in a manner similar to that of a computing language such as FORTRAN or C. CNS modules and procedures have defined sets of input (and output) parameters that are passed into them (or returned) when they are invoked.

Parameters passed into a module or procedure inherit the scope of the calling task file/module, and thus they exhibit a behavior analogous to most computing languages. Symbols defined within a module or procedure are purely local variables.

Library Modules

CNS library modules include space-group information, Gaussian atomic form factors, anomalous scattering components, NMR random-coil chemical shifts, molecular parameter and topology databases, and a

conformational database which contains multi-dimensional probability distributions for preferred rotamers in proteins and nucleic acids.

Task files

Task files consist of CNS language statements and module invocations. The CNS language permits the design and execution of nearly any numerical task in X-ray crystallographic and NMR structure determination using a minimal set of 'hard-wired' functions and routines.

Each task file is divided into two main sections:

-The initial parameter definition: contains the definitions of all CNS parameters which are used in the main body of the task file. It also contains derivatives that specify HTML features, e.g. text comments (indicated by {*...*}), user-modifiable fields (indicated by {= = = >}), and choice boxes (indicated by {+choice:...+}).

-The main body of the task file

Modification of the main body of the task file is not required but can be performed in order to experiment new algorithms.

The task files produce a number of output files, e.g. coordinate, reflection, graphing and analysis files. Comprehensive information about input parameters and results of the task are provided in these output files, so that the majority of the information required to reproduce structure determination is kept within the results. Analysis data is often provided in simple columns and rows of numbers. These data files can be used for graphing. An HTML graphical plotting interface is planned which makes use of these analysis files. In addition list files are often produced that contain a synopsis of the calculations incurred.

HTML Interface

The HTML graphical interface makes use of the HTML form syntax to create a high-level menu-driven environment for CNS. Compact and relatively simple Common Gateway Interface (CGI) conversion scripts are

available that transform a task file into a form page, and the edited form page back into a task file. These conversion scripts are written in the PERL language.

A comprehensive collection of task files are available for crystallographic phasing and refinement, and for NMR structure calculation. New task files can be created or existing ones modified in order to address problems that are not currently met by the distributed collection of task files. The HTML graphical interface thus provides a common interface for distributed and 'personal' CNS task files.

Symbolic target function

One of the key features of CNS is the ability to symbolically define target functions and their first derivatives for crystallographic searches and refinement. A major advantage of definition of the target function and their first derivatives, is tat any arbitrary function of structure-factor arrays can be used. This means that the scope of possible targets is not limited to least-squares targets. Symbolic definition of numerical-integration over unknown variables is also possible. Thus, even complicated maximum-likelihood target functions can be defined using the CNS language. This is particularly valuable at the prototype stage. The standard maximum-likelihood targets are provided through the CNS FORTAN77 code which can be assessed as functions in the CNS language. The availability of internal FORTAN77 subroutines for the most computing-intense target functions and the symbolic definitions involving structure-factor arrays allows for maximal flexibility and efficiency.

Combined maximum-likelihood and Simulated Annealing refinement

CNS has a comprehensive task file for simulated annealing refinement of crystal structures using Cartesian or torsion-angle molecular dynamics. This task file automatically computes a cross-validated σ_A estimate, determines the weighting scheme between the X-ray refinement target function and the geometric energy function, refines a flat bulk solvent model and an overall anisotropic B value of the model by the least-squares minimization, and subsequently refines the atomic positions by simulated annealing. Options are available for

specification of alternate conformations, multi-conformer refinement, and non-crystallographic symmetry. Available target functions include the maximum-likelihood functions MLF,MLI and MLHL. The user can choose between slow-cooling and constant-temperature simulated annealing, and the respective rate of cooling and length of the annealing scheme.

During simulated annealing refinement the model can be significantly improved. Therefore it becomes important to recalculate the cross-validated σ_A error estimates, and the weight between X-ray diffraction target function and the geometric energy function in the course of the refinement. This is important for the maximum-likelihood target functions which depend on the cross-validated σ_A error estimates. In the simulated annealing task file, the recalculation of σ_A values and subsequently the weight for the crystallographic energy term are carried out after initial energy minimization, and also after molecular dynamics simulated annealing.

NMR structure calculation

The NMR structure calculation protocols in CNS consist of four main sections:

data input : includes NOE-derived distances, NOE intensities, torsion-angle restraints, coupling constants, ^{1}H chemical shifts, $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ secondary shifts, dipolar coupling data, and heteronuclear T_1/T_2 ratios.

annealing protocols: the starting points are randomized extended strands corresponding to each disjoint molecular entity or pre-folded structures. The first section of the protocol consists of reading the various data structures. This is followed by an initialization section for statistical analysis of average properties. A constant high-temperature Cartesian or torsion-angle annealing stage follows, followed by a slow-cooling stage with either torsion angle or Cartesian dynamics. Finally an additional Cartesian dynamics cooling stage and a minimization stage follow.

acceptance tests : a number of trials are performed by starting the simulated-annealing calculation with different randomly selected initial atomic velocities.

analysis of all NMR structures: includes analysis of deviations and violations for the various experimental and chemical restraints, which is written out to the header sections of the coordinate file corresponding to the particular trial. The acceptability of the trial is tested and analysis of average properties is carried out.

The whole process begins again using different initial velocities (or coordinates) which in general produces a different result.

REFERENCES

1. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, Shindyalov, L. N., and Bourne, P. E., The Protein Data Bank. *Nuc. Acids Res.* **28**. 235-242 (2000).
2. Billeter, M., Riek, R., Wider, G., Hornemann, S., Glockshuber, R. & Wüthrick, K. Prion protein NMR structure and species barrier for prion diseases. *Proc. Natl. Acad. Sci. USA* **94**, 7281-7285 (1997).
3. Brian J.Bennion & Valerie Daggett. Protein conformation and diagnostic tests: The prion protein. *Clinical Chemistry* 48: 2105-2114 (2002).
4. Brown, P. and Gajdusek, D. C. The human spongiform encephalopathies: kuru, Creutzfeldt-Jakob disease, and the Gerstmann-Straussler-Scheinker syndrome. *Curr. Top. Microbiol. Immunol.* **171**, 1-20 (1991).
5. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L.
6. Calzolai, L. & Zahn, R. Influence of pH on NMR structure and stability of the human prion protein globular domain. *J. Biol. Chem.* **278**, 35592-35596 (2003).
7. Capellari, S., Parchi, P., Russo, C. M., Sanford, J., Sy, M. S., Gambetti, P. & Petersen, R. B. Effect of the E200K mutation on prion protein metabolism. *Amer. J. Pathol.* **157**, 613-622 (2000).
8. Caughey, B. W., Dong, A., Bhat, K. S., Ernst, D., Hayes, S. F. & Caughey, W. S. Secondary structure analysis of the scrapie-associated protein PrP 27-30 in water by infrared spectroscopy. *Biochemistry* **30**, 7672-7680 (1991).
9. Computational Science Education Project ; Mathematical Optimization ; http://csep1.phy.ornl.gov/CSEP/MO/MO.html (1995) (Accessed: January 2005)
10. Crystallography and NMR System: A new software suite for macromolecular structure determination. *Acta Cryst.* **D54**. 905-921 (1998).
11. Cui, F., Jernigan, R. L. & Wu, Z. Refinement of NMR-determined protein structures with database-derived distance constraints. *Submitted* (2004).
12. Cui, F., Jernigan, R. L. & Wu, Z. Enhancement of torsion angle constraints in NMR structure refinement via database-derived distance. *Submitted* (2004).
13. Doreleijers, J. F., Mading S., Maziuk D., Sojourner K., Yin, L., Zhu, J., Makley, J. L., and Ulrich, E. L. BioMagResBank database with sets of experimental NMR constraints corresponding to the structures of over 1400 biomolecules deposited in the Protein Data Bank. *J. Biomol. NMR* **26**. 139-146 (2003).
14. Gossert A. D., Bonjour S., Lysek D. A., Fiorito F., Wuthrich K. Prion protein NMR structures of elk and of mouse/elk hybrids. *Proc Natl Acad Sci USA*. **102**, 646-650 (2005).
15. Goldfarb, L. G., Brown, P., Cervenakova, L. & Gajdusek, D. C. Genetic analysis of Creutzfeldt-Jakob disease and related disorders. *Phil. Trans. Roy. Soc. Lond. B Biol. Sci.* **343**, 379-384 (1994).
16. Haire, L. F., Whyte, S. M., Vasisht, N., Gill, A. C., Verma, C., Dodson, E. J., Dodson, G. G. & Bayley, P. M. The crystal structure of the globular domain of sheep prion protein. *J. Mol. Biol.* **336**, 1175-1183 (2004).
17. Heller, J., Kolbert, A. C., Larsen, R., Ernst, M., Bekker, T., Baldwin, M., Prusiner, S. B., Pines, A., and Wemmer, D. E. Solid-state NMR studies of the prion protein H1 fragment. *Protein Science* **5**, 1655- 1661 (1996).
18. Hornak J.P. The Basics of NMR; Source: http://www.cis.rit.edu/htbooks/nmr (2004). (Accessed: January 2005)
19. Horwich, A. L. & Weissman, J. S. Deadly conformations – protein misfolding in prion disease. *Cell* **89**, 499-510 (1997).
20. Hsiao, K., Meiner, Z., Kahana, E., Cass, C., Kahana, I., Avrahami, D., Scarlato, G., Abramsky, O.,Prusiner, S. B. & Gabizon, R. Mutation of the prion protein in Libyan Jews with Creutzfeldt-Jakob disease. *N. Engl. J. Med.* **324**, 1091-1097 (1991).
21. Kaneko, K., Zulianello, L., Scott, M., Cooper, C. M., Wallace, A. C., James, T. L., Cohen, F. E. & Prusiner, S. B. Evidence for protein X binding to a discontinuous epitope on the cellular prion protein during scrapie prion propagation. *Proc. Natl. Acad. Sci. USA* **94**, 10069-10074 (1997).
22. Knaus, K., Morillas, M., Swietnicki, W., Malone, M., Surewicz, W. K. & Yee, V. C. Crystal structure of the human prion protein reveals a mechanism for oligomerization. *Nat. Struct. Biol.* **8**, 770-774 (2001).

23. Kuszewski, J., Gronenborn, A. M., and Clore, G. M. Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. *Protein Sci.* **5**. 1067-1080 (1996).
24. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291 (1993).
25. Lee I.Y., Prusiner S.B, and Yao H. Complete Genomic Sequence and Analysis of the Prion protein gene region from three mammalian species. *Genome Research* 8: 1022-1037 (1998).
26. Levy Y, and Becker, O. M. Conformational polymorphism of wild-type and mutant prion proteins: Energy landscape analysis. *Proteins* **47(4)**, 458-68 (2002).
27. Meyer, R. K., McKinley, M. P., Bowman, K. A., Braunfeld, M. B., Barry, R. A. & Prusiner, S. B. Separation and properties of cellular and scrapie prion proteins. *Proc. Natl. Acad. Sci. USA* **83**, 2310-2314 (1986).
28. Miyazawa, S. and Jernigan, R. L. Residue-residue potentials with a favourable contact pair term and an unfavourable high packing density term for simulation and threading. *J. Mol. Biol.* **256**. 623-644 (1996).
29. Miyazawa, S. and Jernigan, R. L. Estimation of effective inter-residue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* **18**. 1985, 534-552 (1985).
30. Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. Stereochemical quality of protein structure coordinates. *Proteins* **12**, 345-364 (1992).
31. Oesch, B., Westaway, D., Walchli, M., Mckinley, M. P., Kent, S. B., Aebersold, R., Barry, R. A., Tempst, P., Teplow, D. B. Hood, L. E., *et al.* A cellular gene encodes scrapie PrP 27-30 protein. *Cell* **40**, 735-746 (1985).
32. Pan, K. M., Baldwin, M., Nguyen, J., Gasset, M., Serban, A., Groth, D., Mehlhorn, I., Huang, Z., Fletterick, R. J. Cohen, F. E., *et al.* Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc. Natl. Acad. Sci. USA* **90**, 10962-10966 (1993).
33. Prusiner, S. B. Prions. *Proc. Natl. Acad. Sci. USA* **95**, 13363-83 (1998).
34. Prusiner, S. B. Novel proteinaceous infectious particles cause scrapie. *Science* **216**, 136-144 (1982).
35. Prusiner, S. B. Shattuck lecture — neurodegenerative diseases and prions. *N. Engl. J. Med.* **344**, 1516–152 (2001).
36. Ramachandran, G. N. & Sasisekharan, V. Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283-437 (1968).
37. Rhodes G. Judging the quality of Macromolecular models; University of Southern Maine, http://www.usm.maine.edu/~rhodes/ModQual/#Ramachandran%20diagram (2000) (Accessed: March 2005)
38. Riek, R., Hornemann, S., Wider, G., Billeter, M., Glockshuber, R. & Wüthrich, K. NMR structure of the mouse prion protein domain PrP (121-321). *Nature* **382**, 180-182 (1996).
39. Robert J. Elbourn The prion diseases: a molecular and genetic perspective. Fortunecity (1999); http://www.fortunecity.co.uk/roswell/psychic/24/prionpage/Project.htm#5 (Accessed: October 2004)
40. Rojnuckarin, A. and Subramaniam, S. Knowledge-based potentials for protein structure. *Proteins* **36**. 54-67 (1999).
41. Rudd, P. M., Wormald, M. R., Wing, D. R., Prusiner, S. B., and Dwek, R. A. Prion glycoprotein: structure, dynamics, and roles for the sugars. *Biochemistry* **40**, 3759-3766 (2001).
42. Safar, J., Roller, P. P., Gajdusek, D. C. & Gibbs, C. J. Jr. Thermal stability and conformational transitions of scrapie amyloid (prion) protein correlate with infectivity. *Protein Sci.* **2**, 2206-2216 (1993).
43. Sippl, M. J. and Weitckus, S. Detection of native-like models for amino acid sequence of unknown three-dimensional structure in a database of known protein conformations. *Proteins* **13**. 258-271 (1992).
44. Sippl, M. J. Calculation of conformational ensembles from potentials of mean force, *J. Mol. Biol.* **213**. 859-883 (1990).
45. Stahl, N., Baldwin, M. A., Teplow, D. B., Hood, L., Gibson, B. W., Burlingame, A. L. & Prusiner, S. B. Structural studies of the scrapie prion protein using mass spectrometry and amino acid sequencing. *Biochemistry* **32**, 1991-2002 (1993).
46. Telling, G. C., Scott, M., Hsiao, K. K., Foster, D., Yang, S. L., Torchia, M., Sidle, K. C., Collinge, J., DeArmond, S. J. & Prusiner, S. B. Transmission of Creutzfeldt-Jakob disease from humans to transgenic mice expressing chimeric human-mouse prion protein. *Proc. Natl. Acad. Sci. USA* **91**, 9936-9940 (1994).

47. Telling, G. C., Scott, M., Mastrianni, J., Gabizon, R., Torchia, M., Cohen, F. E., DeArmond, S. J. & Prusiner, S. B. Prion propagation in mice expressing human and chimeric PrP transgenes implicates that interaction of cellular PrP with another protein. *Cell* **83**, 79-90 (1995).
48. Wall, M. E., Subramaniam, S., and Phillips, Jr. G. N. Protein Structure Determination Using a Database of Inter-Atomic Distance Probabilities. *Protein Sci.* **8.** 2720-2727 (1999).
49. Wu.Z Department of Mathematics; http://orion.math.iastate.edu/wu/ (2004). (Accessed 28 June 2005)
50. Zahn, R., Liu, A., Lührs, T., Riek, R., von Schroetter, C., Garcia, F. L., Billeter, M., Calzolai, L., Wider, G. & Wüthrich, K. NMR solution structure of the human prion protein. *Proc. Natl. Acad. Sci. USA* **97**, 145-150 (2000).
51. Zhang, Y., Swietnicki, W., Zagorski, M. G., Surewicz, W. K. & Sönnichsen, F. D. Solution structure of the E200K variant of human prion protein. *J. Biol. Chem.* **275**, 33650-33654 (2000).

ACKNOWLEDGEMENT

With a deep sense of gratitude, I wish to express my sincere thanks to my advisor, Dr. Zhijun Wu, for his immense help in planning and executing the works in time. The confidence and dynamism with which Dr. Wu guided my research work requires no elaboration. His valuable suggestions and words during the course of my research are greatly acknowledged and cherished.

My sincere thanks are due to the Department of Mathematics, Iowa State University, for providing me the opportunity as an International Student, to learn and obtain further education under the guidance of distinguished mathematicians.

I would also like to thank the members of my research group for the help extended to me when I approached them during the length of my study and the valuable discussion that I have had with them. The cooperation I received from other faculty members of this department is gratefully acknowledged. I will be failing in my duty if I do not mention the administrative staff of this department for their timely help.

I would also like to thank my parents, who taught me the value of hard work by their own example. I would like to share this moment of happiness with my family. They rendered me enormous support during the whole tenure of my research. The encouragement and motivation that was given to me to carry out my research work by them has granted me the success, for which I am forever indebted.

Finally, I would like to thank all whose direct/indirect support helped me complete my thesis in time.